

Whither high-energy physics?

Last week's meeting of the CERN council has sharpened the dilemma of those who suck their teeth about the cost of high-energy physics. The time has come to plan for wider collaboration.

It is a great relief that the British government told the CERN council at Geneva last week (see p. 681) that it intends to stay a member of the organization, but that CERN (the European Organisation for Nuclear Research) must find ways of cheapening both its total cost and the cost of membership. It would have been a great misfortune for Britain and, to a lesser extent, for CERN if Britain had withdrawn. The hope now must be that the legitimate pursuit of economies stops short of damaging CERN's scientific programme. The ideal should be that the whole contentious episode should be concluded at the special meeting of the council that is due to take place in two months' time.

Meanwhile, in the administration of British civil science, the need is now more urgent than ever that the budget of the Science and Engineering Research Council should in future be insulated from the fluctuations of the cost of membership arising from fluctuations of the relative cost of sterling and Swiss francs. If, on the grounds of general financial policy, the British government manipulates the value of sterling so as to retain some relationship with the US dollar, but the Swiss government does not, can it be said to be equitable that there should then be less money to spend on the general support of academic research in Britain?

That is a domestic issue for the British, but the report of the committee under Professor Anatole Abragam, set up to identify potential economies in CERN's high-energy physics, provides other members of CERN with an interesting opportunity. The report is both supportive and critical of CERN. Abragam and his colleagues are full-throated in their applause for CERN's achievement as a high-energy physics laboratory, and for its plans for the years ahead, which is only right and proper. Their criticisms are of a managerial character: the working committee responsible for CERN's policy between council meetings has grown too big; liaison with industry and even scientific users should be improved; the costs of informal arrangements with high-energy physicists outside the CERN membership are often unrequited; the permanent staff includes 300 people who will not be needed in the 1990s; and too great a proportion (86 per cent) of the full-time staff at Geneva (3,500 in total) enjoy indefinite employment contracts.

Wider interests

To be fair to CERN's present generation of managers, most of Abragam's complaints are implicitly levelled at previous decisions (sometimes taken by default) of the governing council, which is also responsible for the prevalence of the salary supplements enjoyed by those who work at CERN (who are also exempt from income taxes). But it is in the wider interest, not just that of keeping the British government quiet, that Abragam's recommendations should be followed energetically. As the committee report puts it, CERN exists to serve the European high-energy physics community, without which it would have no independent existence, and must therefore be an integral part of it. This is the spirit in which the committee would have CERN shed 300 people sooner rather than later, and would

reduce the proportion of people on indefinite contracts to below a half. The report also has sensible things to say about the relationship between the laboratory and its social environment, that of Switzerland and France, which could lead to further reductions of employment costs. But with the best will in the world, the savings in the annual budget will amount to only 10 per cent or so, less than the British government had been hoping for.

Meanwhile, plans are being drawn for the further development of CERN. Last week's council meeting was presented not merely with the Abragam report but with a document from CERN's own committee under Carlo Rubbia set up to work out plans for the distant future. Like much else of what CERN does, the forward plan is at once imaginative and convincing.

Proposals

The version of LEP (for Large Electron-Positron collider) now being built will arrange for the collision of 50-GeV electrons and positrons, but it is intended that the energy of each beam should be increased to 100 GeV by the use of more powerful magnetic fields. What the Rubbia committee argues is that, even then, much interesting physics will be out of range, and that crucial tests of current theories of particle physics will come only at a collision energy two orders of magnitude greater, at 10 TeV or thereabouts. So the committee advocates two developments — the use of the LEP tunnel (27 km in circumference) to arrange for the collision of protons and electrons and the construction of a novel electron-collider based on two linear accelerators accelerating particles to more than 1 TeV. (Such a scheme has also been proposed by the Nuclear Physics Institute at Novosibirsk, and may yet be built at Serpukhov.)

All this is fine and dandy. Implemented, one or other of these plans would help to keep European high-energy physics on a par with that in the United States and the Soviet Union (not to mention whatever may be happening in Japan in the year 2000). The obvious difficulty is that the costs, estimated only roughly for last week's meeting, are bound to be very great, while full utilization of the new machines (if built) would require the maintenance of the European high-energy physics community at roughly its present size. Is that not a recipe for future trouble, not necessarily with the British then but with some other member state? That is why it would make sense if, in the wake of the Abragam report, CERN were to take the initiative with proposals for fuller (and no doubt more effective) collaboration on the construction of the next generation of accelerators. Everybody in the field acknowledges that the time must come when particle accelerators are built on a thoroughly international basis — then, however, follows latter-day Saint Augustine with the predictable reservation "... but not yet". As it happens, CERN has unique experience of international collaboration at work, and is best placed to propose to a wider range of partners means by which fuller collaboration might be organized. Nobody would blame it if others' chauvinism caused it to fail, but CERN will have only itself to blame for the consequences of not trying. □



Privatization for NIH?

Floating off intramural research at NIH is not altogether a bad idea.

The infectious fashion for turning over public assets to private ownership, inelegantly called privatization by the British government (which claims to have invented the notion), seems now to have touched even the US National Institutes of Health (NIH). By all accounts (or at least by that of the *New York Times* last week), the Office of Management and Budget (OMB) has been working on a proposal for turning the part of NIH that carries out research in-house into a private institution, a kind of research university under independent management. The first reactions to the news, in Congress and elsewhere, have been expressions of alarm; why tamper with success? But the proposal deserves more careful attention than a reaffirmation of the comforts of the *status quo*. The acid test of OMB's proposal should be a considered evaluation of the benefits and disadvantages of radical change.

The organization is at present two things — a biomedical research organization and, by delegation, the US government's chief grant-making agency for biomedical research. For the past several decades, both activities have been helped to grow by the good will of a well-meaning Congress, which habitually appropriates more money for NIH than the administration requests. But congressional indulgence for NIH is alloyed with an inclination to interfere, both in the management of NIH's in-house research and in their structure, with repeated special pleading on behalf of diseases newly drawn to the attention of interested congressmen. At various times, moves of this kind have threatened seriously to skew the balance of NIH's in-house activity and to cramp the institutes' flexibility. The *status quo* abounds with danger, which is another reason why those concerned for the welfare of NIH should not too quickly scorn OMB's initiative.

At the same time, the considerations that have apparently prompted the proposals are not in themselves a sufficient justification for such a radical change. NIH are said to be concerned that their status as a part of the public service, which requires that in-house researchers earn government salaries, will lead to the loss of able people to other institutions able to offer more. Dr Robert Gallo's flirtation over several months with universities and commercial companies is said to have brought the issue to a head. But there is no good reason why NIH's constitution should be changed simply to keep able people within its walls. Might it not, on the contrary, be that an important part of NIH's social function is to populate university and other research laboratories with able leaders? If Gallo ends up at Johns Hopkins or Duke universities, or even somewhere else, NIH's loss will be some other institution's gain.

It matters more that, if NIH's in-house research became the responsibility of a private corporation, able to decide for itself on the balance (and even the location) of its research, Congress might get even better value for the \$600 million plus now spent of these activities (just over 10 per cent of the total budget). Moreover, in-house NIH would then be able to play a fuller part in the educational life of the United States than they do at present. The other side of that coin is, however, plain. Would an independent research organization of such great size continue to enjoy the favours of the US Congress if it were competing for public funds with other private institutions, universities for example, to which congressmen may have conflicting constituency loyalties? If, on the other hand, private NIH would have to raise their own funds from the grant-making part of the present organization and from private sources, would their future seem secure enough to satisfy even the researchers who now envy the higher salaries obtainable elsewhere? Left to themselves, NIH would probably settle for the *status quo*. Much will depend, in the months ahead, on the extent to which OMB persuades itself

(and others) that making in-house NIH compete with others for research support would markedly enhance the quality of research. Experience elsewhere, in Britain for example, shows that a little uncertainty can be beneficial but that too much of it is damagingly distracting. □

Research for industry

The British government wants industry to pay for public development, but should go carefully.

SUCCESSIVE British governments have wrung their hands over the reluctance of British companies to invest a greater share of their resources in research and development. The present government is no exception, but its exhortations have not been conspicuously more effective than those of others. That, together with the wish to free resources for basic research, explain why it has now embarked on a scheme for shifting from its own budget the cost of civil research judged to be so much in the nature of development, rather than research, that it could reasonably be paid for by its eventual users.

The Cabinet Office is in the middle of a review of government spending on what it calls "near-market" research, among which the Department of Energy's support of nuclear power developments is conspicuous, but which touches all kinds of government activities in fields as different as agriculture and information technology. The underlying objective seems to be not merely economy but the application of an extension of the Rothschild principle of the 1970s: research and development with industrial application is likely to be the more effective when the putative beneficiaries have to pay for it.

The most obvious danger stems from the simple truth that industrially orientated research is likely to be most effective not merely when the beneficiaries pay for it, but when they actually carry it out themselves, privately (for the sake of commercial secrecy) and in circumstances in which researchers and salesmen rub shoulders easily. Most companies faced with the prospect of having to pay for research carried out in public laboratories ostensibly on their behalf will prefer to do the work themselves, even at much greater cost. So enforcement of the extended Rothschild principle may indeed have the effect of generating a greater volume of in-house research by forward-looking companies. But the more immediate consequence may be further to shrink the volume of work at government research laboratories, with more putting-out to grass of able researchers and the like.

Moreover, there are some fields in which the principle is plainly inapplicable. Agriculture is one; it would be inefficient and probably inequitable if the government should seek to recover the costs of applied agricultural research in some of the obviously tempting ways, a levy on farm produce for example, while making food manufacturers and retailers pay for it will break the crucial link between research and its beneficiaries.

At the other end of the spectrum, nuclear power is another field in which continued public investment is, on present policies, essential. Britain, like most other producers of nuclear power except the United States, reprocesses irradiated fuel so as to extract plutonium and the other actinides from irradiated uranium, both for the sake of the extra energy that may be wrung from them and because incinerating long-lived fissile material in fast reactors is the safest way of disposing of it. The result is that Britain, like other European states, spends public money on the development of fast reactors. But is that not a "near-market" activity whose cost could be spun off to the Central Electricity Generating Board? That is what the government will argue. The other side of that argument is that the government, which has endorsed the policy of reprocessing uranium, needs itself to be informed by a fast-breeder research programme if only to keep intelligently in mind the wisdom of its policy. What all this implies is that there are probably fewer opportunities for wise economies than appear at first sight. □

Britain stays in CERN but wants "extensive savings"

- Relief among scientists as pressure lifted
- UK government persuaded by review

London

THE British government seems to have been persuaded of the merits of remaining a member of the Geneva-based European Organization for Nuclear Research (CERN). The British delegate told last week's meeting of CERN council that "the UK wishes to remain a full and active member of CERN provided that a sound base for doing so can be established". The announcement was greeted with relief by British scientists at CERN; it had been feared that Britain would give provisional notice to quit unless an acceptable level of saving could be identified. In the event, Britain was content to place weight on the findings of the CERN review committee, chaired by Anatole Abragam.

Britain's official statement to the council read: "Council must respond positively to the Abragam report and must not allow the recommendations to be watered down — the situation is far too serious for that". It continued: "It will be difficult for the UK to remain a member unless major savings can be made in the cost to existing member states". Britain urged the council to look at ways of widening membership

of CERN to include non-European states. It is still unclear whether the Abragam recommendations will produce the sort of savings that Britain deems acceptable.

The final report of the review committee differs little from the unpublished interim report, presented to member state governments last June (see *Nature* 327, 648; 1987). Abragam remains adamant that CERN's science must remain untouched. Instead, "extensive budget savings" could be achieved principally through improved management efficiency and reduced staffing levels. A reduction in the scientific programme would "bring only relatively small savings out of proportion with their drawbacks".

Personnel costs at CERN account for SF377 million in 1987 (the annual materials budget is SF395 million), covering the salaries of 3,340 staff members, 91 per cent of whom have indefinite contracts. There are more than 900 service-contract personnel, many of whom work permanently at CERN, and some 4,000 registered outside users. The report points out that while other organizations employ most staff on indefinite contracts, CERN has to provide a pension fund.

Abragam recommends a programme of early departures, with the loss of 300 staff through voluntary and, if necessary, compulsory redundancy in the next three years, together with natural departures at a rate of 60 per year. The cost of termination indemnities and compensation to the pension fund would be SF115 million, which would need to be borrowed and repaid by 1993.

In addition, Abragam would reduce from the present 86 per cent to 50 per cent the proportion of indefinite contracts for 'category A' employees (including physicists, engineers, technical staff and higher administrative staff). Abragam would also like to see examined the possibility that support staff might be employed by CERN subsidiaries under Swiss or French law, abolishing their present international-employee status. CERN management is thought to be unhappy with this. The report advocates new personnel practices, including performance assessment.

CERN management is understood to be in agreement with much of what Abragam recommends, and has started to implement several of the suggestions. But some resentment has been engendered by what is seen as a subjectively aggressive tone to the report.

Simon Hadlington

US budget down under the wire

Washington

THE silly season has arrived for the US budget process. The government officially ran out of money at the end of the day on Friday, 18 December, but Congress held an extraordinary Sunday session implementing a one-day extension to cover spending until midnight on Monday, 21 December. As *Nature* goes to press it is anybody's guess whether a final agreement on the budget can be reached before the current extension expires. If Congress fails to adopt an extension or to pass the total 1988 spending package, the government will technically have to shut down.

Nearly all government spending is contained in a \$600,000 million appropriations bill. Negotiators from the House of Representatives and the Senate have been meeting for weeks to resolve spending differences between the two chambers. As the budget process marches inexorably to completion, last-minute changes for government agencies are not only possible, but likely.

Even when the Congress decides on the numbers, the language in the bills, still to be written, will affect how the money may be spent. The proposed doubling of the National Science Foundation (NSF) budget does not appear to be occurring as rapidly as originally proposed. As of last week, NSF's spending figure was \$1,717 million, \$94 million more than 1987, but \$176 million less than requested. It is still uncertain how these new figures will affect spending on the science and technology centres proposed by NSF, hailed as crucial for US economic competitiveness.

Also last week, conferees agreed to spend \$425 million on the space station, some \$342 million less than NASA had been seeking. The spending is apparently authorized in two parts; \$200 million now and \$225 million after June 1988.

Congress is in no mood to commit to major new spending projects, as finding politically acceptable places to cut the budget is nearly impossible. Joseph Palca

Research spending in the Netherlands

Waalre, The Netherlands

REVISED figures from the Dutch Ministry of Education and Science show that investment in research and development during 1985 and 1986 was far greater than previously thought. In contrast to earlier estimates of a 7.5 per cent growth, the growth achieved was 16.7 per cent for 1985 and 13.4 per cent for 1986. Growth estimates for 1987 and 1988 have consequently been revised from 3.4 to 5 per cent.

The total national investment in research in 1987 was about Dfl10,100 million (about £3,000 million), about 2.34 per cent of the gross national product (GNP), compared with a previously published figure of Dfl 9,300 million, 2.15 per cent of GNP.

The new projection for science spending in 1988 is Dfl 10,400 million, compared with a previous estimate of Dfl 9,350 million. Based on the new figures, industry's investment in research in 1987 was Dfl 5,730 million, and that of government was Dfl 4,183 million, of which the universities received Dfl 1,630 million, Dfl 70 million less than in the previous three years.

Casper Schuurin

New Pasteur director

Paris

THE Pasteur Institute in Paris last week appointed a new director. Dr Maxime Schwartz, aged 47, a molecular biologist, will replace the outgoing director, Raymond Dedonder, whose term of office has expired. Schwartz joined the Pasteur in 1963 and in 1972 became head of the molecular genetics unit. In 1985 he was elected deputy director of the institute, having been a member of the scientific council and head of the molecular biology department for several years.

P.C.

Congress goes for Nevada as site for nuclear waste storage

Washington

Thirty years of debate and uncertainty have been brought to an abrupt end with a sudden congressional decision to name Yucca Mountain, Nevada, as the site of the United States' first repository for high-level nuclear waste. The decision, one of the many last-minute deals made in Senate/House of Representatives conference committees as time ran out for agreement on the 1988 budget (see page 681), throws out the scientific site-selection process laid out in the 1982 Nuclear Waste Disposal Act. And although the legislation will still have to be approved by both Senate and House, political expediency seems certain to win the day.

State officials in Nevada reacted with fury as they found the dump had been forced on them. Governor Richard H. Bryan attacked Congress's decision as a "legislative atrocity" that "blatantly rejects the laws of the land" and promised a "nuclear nightmare for Congress and the utility industry". The state will fight in the courts and "through whatever other avenues are needed".

The selection process abandoned in the new legislation had required the Department of Energy to prepare a short-list of six sites — three west of the Mississippi and three east — each stable enough to store radioactive waste for 10,000 years.

A site would then have been chosen in the west and, some years later, another in the east. But the process proved unworkable, despite the years of debate that went into its making. In May 1987, Energy Secretary John Herrington tried to take a short-cut by announcing three sites for investigation in the west — Hanford, Washington; Deaf Smith, Texas; and Yucca Mountain, Nevada — and giving up the search for sites in the east. But the resulting protest from westerners quickly led to demands to restart the whole process. At the same time it became clear, given the level of local opposition everywhere, that it might take decades to complete the assessment of even three sites. Meanwhile, nuclear waste would continue to pile up in temporary storage.

Earlier this year, J. Bennett Johnston, chairman of the Senate Energy Committee, attempted to find a way out by proposing that a \$100 million-a-year incentive be paid to any state willing to open the first high-level nuclear-waste dump. But the candidate states did not take the bait.

Now Nevada seems likely to be saddled with the dump. Local opponents remain convinced, however, that, in Governor Bryan's words, "scientific and technical concerns ignored on Capitol Hill will eventually disqualify Yucca Mountain".

The great advantage of Yucca Mountain, a 1,500-foot-high ridge of volcanic rock, is that it lies well above the water table. A shallow storage area, 1,200 feet below ground, would seem to permit easy access while avoiding the possibility that groundwater could leach radioactive wastes from storage containers. And Yucca Mountain is in one corner of the Nevada Nuclear Test Site and far from human habitation.

But state geologists argue that the site lies near an active fault and that a major earthquake might split it open. There is also evidence that a volcanic eruption might occur. More immediately, the local tourist industry might suffer when visitors find they have to share roads with trucks bearing nuclear waste.

All these fears will have to be addressed in the ensuing site investigation. And, if they turn out to be well founded, then the whole process will have to begin again at a

new site with a delay of many years. That risk, inevitable in investigating only one site, is one the Department of Energy says it is prepared to take because of the savings that will be made should Yucca Mountain be the right choice.

That site selection is no easy matter is clear from revelations last week concerning the United States' first permanent underground nuclear storage site, for lower-level wastes, near Carlsbad, New Mexico. Seven hundred million dollars have already been spent on a vast network of tunnels and storage rooms burrowed 2,000 feet into thick salt deposits. Back in 1957, the National Academy of Sciences recommended salt deposits as stable repositories because "no water can pass through salt". But, as University of New Mexico geologist Roger Y. Anderson put it, "the area was selected in haste and there have been plenty of geological surprises". The latest surprise is that the walls are weeping and there is a risk that drums of plutonium-contaminated waste from the manufacture of nuclear weapons may end up sitting in pools of corrosive brine.

Alun Anderson

Massachusetts adopts a model radwaste law . . . at long last

Boston

AFTER more than five years of negotiation and controversy, the state of Massachusetts recently enacted new regulations for the monitoring and storage of low-level radioactive waste. The bill, signed by the governor this month, has been hailed by industry and environmental observers alike as one of the most coherent and comprehensive in the country.

Due in part to the large number of hospitals, research laboratories and commercial companies specializing in biotechnology or the production of radiopharmaceutical products, Massachusetts ranks fifth among the states as a producer of low-level radioactive material, disposing of some 120,000 cubic feet of wastes and 110,000 curies annually. Until now, the state has handled these wastes under a series of essentially *ad hoc* regulations and has shipped them to one of three low-level waste disposal facilities in the United States: Hanford, Washington; Barnwell, South Carolina; or Beatty, Nevada.

Notably, the Massachusetts law mandates a new state agency intended to consolidate legal and technical expertise in radioactive waste management, which would monitor the long-term storage of these substances. In addition, although the bill does not specify the site or the technology of the state's ultimate disposal facility, it sets out a timetable and strict guidelines for its development.

Among these stipulations is an explicit

rejection of the one method of disposal currently licensed by the Nuclear Regulatory Commission: shallow land burial. Instead, the bill says that if an in-state facility is ultimately required, it must adopt an "active management technology", one capable of continuously monitoring the stored wastes, and allowing the wastes to be repackaged should leaks or deterioration occur.

Also included in the bill are a new classification system that discriminates between the wastes by a weighted assessment of both toxicity and longevity, and strong state incentives for source reduction by the major waste producers.

In the wake of rising disposal costs and continued threats since the early 1980s that future access to these disposal facilities would be denied, the pressure has been on many states to promulgate long-term management plans. Rhode Island has already adopted a version of the Massachusetts bill, and other states including New Jersey and Maryland are reported to be considering the plan.

Steven Roop, the state's assistant secretary for waste management policy, commended the legislative commission for "toiling over the entire range of issues involved. But he and others expressed frustration at the long process involved, especially since an almost identical form of the bill was endorsed unanimously by a legislative committee in early 1985.

Seth Shulman

Lost Himalayan river found in Rajasthan

New Delhi

A COMBINATION of modern technology and clues from ancient Hindu scriptures has led to the discovery of a lost Himalayan river below the desert sands of Rajasthan in India.

Dr Bimal Ghose, a geomorphologist at the Central Arid Zone Research Institute in Jodhpur, has traced the buried channels of the Saraswati River and says the discovery of the subterranean river may provide a solution for bringing the once lush and green Thar Desert back to life.

Using maps supplied by Ghose, government agencies have already sunk 36 tube-wells along the buried course of the Saraswati in the district of Jaisalmer in the heart of the Thar Desert.

Surprisingly, each of the wells is producing between 2,000 and 40,000 litres of water per hour. There is so much water flowing below the desert, Ghose says, "that if all the tube-wells are operated simultaneously, this part of the desert can be flooded".

Ghose launched his hunt for the Saraswati more than a decade ago after seeing references to the river in two ancient Hindu manuscripts, the Rig Veda and the Mahabharata (3000 BC).

Along with the Indus and the Ganges Rivers, the scriptures made references to two other north Indian rivers originating in the Himalayas, the Saraswati and the Apaya. The Saraswati has now been traced, while the search for the lost Apaya continues.

In mapping the buried course of the Saraswati, Ghose cross-checked the clues from the scriptures with ground truths, aerial photographs and land imageries. Ghose says it is now established that the Saraswati, rising from the Shiwalik Himalayas, once flowed through the Haryana and Rajasthan states into the now-dry district of Gujarat. The Thar Desert appears to have been formed by the Saraswati River changing its course westward and eventually becoming subterranean about 1,000 years ago, which alters the previous conception as to how the desert came into existence.

Although the Saraswati River is dead, Ghose says, its buried channels conduct the water formed by the melting snow of the Himalayas. These buried channels, being perennial, are dependable sources of ground water that could be used to solve permanently the drinking water problem in the desert. Ghose is now trying to delineate the buried deltas of the Saraswati for investigation of the existence of crude oil.

K.S. Jayaraman

Human Frontiers Program seeks international help

Tokyo

JAPAN'S Human Frontiers Program is back in the news again. After being advertised as a major international effort in basic bioscience research, into which Japan might pour ¥1 million million (\$8,000 million) over twenty years, the programme has spent almost two years searching for an identity. Last week more than 30 scientists and science-related government officials from the seven Western summit nations (Japan, France, West Germany, Italy, the United Kingdom and the United States) gathered in Tokyo to try to give the programme some shape. But it is still not at all clear what financial backing will be available.

The programme emerged from the Ministry of International Trade and



Takeshita — commitment to programme?

Industry (MITI) early last year, partly in response to Western criticism that Japan makes too small a contribution to the world pool of basic scientific research (see *Nature* 320, 296; 1986). Its budget, although not exceptionally large by US standards (the US National Institutes of Health spend \$5,000 million a year), would have marked a major advance in support for basic research in Japan.

From the outset, the programme has been driven by ministry officials and has lacked clear scientific goals. Just before the Venice summit where the programme was discussed, former Prime Minister Yasuhiro Nakasone criticized its vagueness.

Scientists in MITI's feasibility study committee made various suggestions before the summit, for example that it should involve research on AIDS (acquired immune deficiency syndrome) and sequencing of the human genome, but no final decision has been made.

A decision on the budget is also lacking. Less than \$2 million has been assigned for a feasibility study which began in December last year and will continue until April next year.

Prospects for 1988–89 are little better; MITI and the Science and Technology Agency (which joined the programme late

last year) have requested a combined total of only about ¥500 million (\$4 million) for a 'pilot' programme in fiscal 1988.

None of this seems to have diluted the effusive enthusiasm of the delegates at last week's meeting. Dr Benno Hess, vice-president of the Max Planck Society, sees the programme as a unique opportunity to form a trilateral research programme between Europe, the United States and Japan. And David Noble, administrative secretary of Britain's Medical Research Council, said it was a "well organized meeting".

In MITI's "tentative" proposal it is suggested that the programme will consist of research grants, fellowships, workshops and the "promotion of research and development of related technologies" (such as DNA sequence analysers and non-invasive body scanners).

A board of trustees from the summit nations will determine the size of budgets, a secretariat will provide administration, and a governing council of 20 distinguished scientists drawn from all participating nations will determine research areas, allocate funds and select outstanding scientists to review applications for grants and fellowships. The proposal also repeatedly emphasizes that the programme should support "young" scientists — a new idea in Japan where the government provides only a few hundred post-doctoral fellowships.

The "tentative" conclusion of the meeting was that emphasis should go to research which cannot be carried out by a single nation. But it was recognized that it might be difficult to organize an international team consisting solely of young researchers — senior scientists might be necessary as organizers.

Research subjects and partners were thought to be best chosen at the international workshops of the programme.

But all discussion will prove superfluous if Japan's Ministry of Finance does not put substantial funds into the programme — and it will now be at least 1989 before it becomes clear where the ministry stands. MITI officials insist that Nakasone's retirement will not take the steam out of the programme. But Nakasone has long been a strong supporter of research in the biosciences and has pushed for the 'internationalization' of Japan. In contrast, his successor, Prime Minister Noboru Takeshita, seems more concerned with domestic politics than with international research in bioscience. And unless the programme gets strong political backing inside and outside Japan it will never achieve the levels of funds first considered.

David Swinbanks

US-Soviet dispute persists over alleged psychiatric practices

Washington & London

US officials have expressed surprise and consternation over remarks made by Soviet Minister of Health Evgenii Chazov about US attitudes towards Soviet psychiatry reported in last week's *Nature* (see 330, 595; 1987). In a recent speech, Chazov said the US Department of Health and Human Services had rejected an offer to form a joint commission to investigate alleged misuse of psychiatry because US psychiatrists were not interested in Soviet psychiatry. But a US Health Department official emphasized that only the specific commission proposed by Chazov was rejected, and that the United States is committed to resolving issues that have prevented scientific exchanges in psychiatry between the two countries.

The US government refuses to allow formal scientific exchanges between US and Soviet psychiatrists to take place because of concern about Soviet psychiatric practices. Particularly upsetting to the US government are practices at the Serbskii Institute of Forensic Psychiatry. Dr Aleksandr Snezhnevskii, former head of the institute, has described all mental illness as a form of schizophrenia, and political dissent as a form of "creeping schizophrenia", a condition unknown to psychiatrists elsewhere.

In 1983, the All Union Society of Psychiatrists and Neuropathologists of the Soviet Union resigned from the World Psychiatric Association (WPA) as WPA

was preparing to expel the society for condoning abusive psychiatric practices. The American Psychiatric Association (APA) will not allow official contact with the All Union Society unless it includes "the opportunity for discussion of abuse of psychiatry" in the Soviet Union, making formal collaboration effectively impossible. Exchanges between individuals may still occur.

Lately, the Soviet media — notably the Communist youth paper, *Komsomolskaya Pravda* — have begun to investigate cases of workers compulsorily hospitalized in psychiatric units, allegedly for being "troublemakers". This has given the Soviet health establishment a renewed interest in showing that psychiatric abuses do not occur. In April, Health Minister Chazov approached Ambassador Richard Schifter, US assistant secretary of state for human rights and humanitarian affairs, asking Schifter to visit psychiatric hospitals in the Soviet Union. But Schifter declined, suggesting instead that a delegation from the APA go.

In August, Ellen Mercer, director of APA's office of international affairs, wrote to the Soviet embassy in Washington, requesting a meeting to discuss a visit by an APA panel. Mercer says the embassy has yet to reply. She says APA's proposed delegation would resolve the question of abuses. The APA asks that a delegation be allowed to visit patients of its choosing, to examine patients in the

hospitals where they are treated and to have family members present during examinations. In addition, the delegation seeks to have its own interpreters, and to have patients taken off their "medications or other treatments designed to alter their behaviour" for the period of the delegation's visit.

One likely reason for the Soviet's silence on the APA's offer is animosity towards the organization. It played a major role in factors leading to the Soviet Union's withdrawal from the WPA.

The Soviet Union approached the Department of Health and Human Services to see if exchanges in psychiatric research could be worked out under the terms of the bilateral health agreement currently in place between the two countries. But following consultation with the State Department the request was refused because of the unresolved issue of abuses, especially at the Serbskii Institute. Harold Thompson, deputy director of the office of international health, says permitting exchanges would have the effect of giving official US government approval for Soviet psychiatry. That, he argues, should be the job of the private sector, in this case the APA.

With formal channels unwilling or unable to provide the sort of delegation sought by the Soviet Union, Chazov has tried yet another route. He is co-chairman of the International Physicians Movement against Nuclear War, for which role he shared the Nobel peace prize with American Bernard Lown in 1985. Chazov suggested to Lown that a group of distinguished US psychiatrists could visit the Soviet Union. Lown passed on the request to Lester Grinspoon, associate professor of psychiatry at Harvard Medical School, who wrote to Chazov on 29 October making essentially the same proposal as the APA. Grinspoon says his group would not represent APA, and thus might enable the Soviet Union to save face by not having to deal with the APA. But he also reminded Chazov that any report his group might write would have more influence if it were supported by the APA. Grinspoon proposes publishing both his delegation's report and a Soviet response to it.

Grinspoon has also not yet received a response. The subject of Soviet psychiatric practices came up at the summit meeting between President Ronald Reagan and General Secretary Mikhail Gorbachev earlier this month in Washington. Both governments reiterated their interest in the issue, but the United States continues to maintain that the APA should conduct such an investigation. Grinspoon hopes that his private initiative will ultimately be acceptable to all sides, and says he expects an answer from Chazov shortly.

Joseph Palca & Vera Rich

Supercomputer compromise still unsettled

Bangalore & Washington

A year-long controversy over the supply of a supercomputer to India from the United States has ended in compromise. Although the terms of the agreement are very general, it seems that India will be offered a Cray XMP-14 supercomputer, rather than the more advanced XMP-24 model which had been requested. But many Indian scientists, concerned that the Cray is soon to go out of production, are backing the purchase of a supercomputer from Control Data Corporation.

According to diplomatic sources in Washington, during the recent visit to the United States of the Indian Prime Minister, Mr Rajiv Gandhi, the United States had assured him of the upgrading of the Cray XMP-14 into the XMP-24 in the foreseeable future. The United States had also agreed to transfer the latest computer software to India.

Meanwhile, Mr Bruce Smart, the United States under secretary, has revealed that India has expressed its interest in another

supercomputer. Control Data Corporation has reached an agreement with India and the State Department to ship up to one single processor from its latest generation of supercomputers including the E series models 1032, 1064, 1096 and 1128 which have memory sizes of 64 or 27 megawords, or the ETA/10T or ETA/10Q models made by ETA Systems Inc. The Bangalore-based Indian Institute of Science, which wants a supercomputer to aid in its monsoon forecasting research programme, may well choose a Control Data Corporation model in preference to the Cray model.

Many well-informed Indians feel that the US willingness to export the Cray is due not to generosity but to the fact that there are no buyers for the XMP series as a consequence of the Cray Corporation's decision to scrap production of the series. And many users of this series, including the British Department of Energy, are now worried about future servicing and maintenance of their system.

Radhakrishna Rao &
Elizabeth Bragunier

Atomic Energy Agency warns of imminent budget crisis

Vienna

DESPITE praise from leaders the world over for its nuclear safety efforts, the International Atomic Energy Agency (IAEA) is facing a financial crisis. The root of the problem is delinquent payments from member nations, not least of them the United States.

Hans Blix, director general of the IAEA, calls it "an absurdity" that the agency, praised for its efficiency and described by US President Ronald Reagan as "a model of effective international co-operation", should have a budget shortfall in 1987 due to US tardiness.

The crunch comes at a critical time for the agency. Its burden of operational safety visits to nuclear power plants has increased from two to twelve a year in the wake of the Chernobyl accident. Even steady payments will not be enough in the near future. "There will be a new enrichment plant coming on line in each of the next two years", says Blix. "We cannot maintain the level of quality of our safety inspections" when the budget is at zero growth. The agency will receive \$137 millions, in 1988, 0.6 per cent more than 1987.

Blix puts most of the blame for the agency's financial woes on the United States — which still owes \$21 million for 1987 — though he makes it clear that it is not the only country that pays late. The IAEA charter calls for each member to pay its dues in June, a deadline which "everybody ignores", says Blix. The IAEA has muddled through the past few years using early payments from some states and relying on a cash surplus that has now evaporated.

The difficulty with the United States results primarily from the bad reputation of international organizations in the

United States Congress, says Blix. Some groups, such as the United Nations Industrial Development Organization, have been hit harder than others. The IAEA has been affected despite having "friends" in the United States Congress and despite its value to a world made nervous by the Chernobyl catastrophe.

"The conclusion after Chernobyl", said Blix, "was that no accident resulting in radioactive release would be tolerable.



Blix — blaming US for money troubles

The member countries came to us to enforce this. But now they seem to be saying that this was only a temporary situation — that we can go back to pre-1986 safety levels. Our view is that if you want a better safety effort, you have to pay for it." The IAEA does not have a very big stick with which to threaten late payers. The worst that can happen is that a member state two years in arrears loses its vote at meetings. Few countries have been deprived of their votes in the 30-year history of the agency.

Steven Dickman

Think-tank set up for British parliament

London

SENIOR Conservative backbench Members of Parliament (MPs) are setting up an independent body to inform parliamentarians on scientific issues, supplementing inadequate information provided by the British government. The Science and Technology Information Service for Parliament (STISP) will advise MPs on matters likely to be debated in Parliament.

STISP is the brainchild of Sir Trevor Skeet, a Conservative MP and chairman of the all-party Parliamentary and Scientific Committee, a body comprising more than 200 MPs, peers and representatives from industry and research institutions.

In the absence of public money to support the venture, STISP will be supported

by a charitable trust with funds from universities, research institutions and industry as well as members of the all-party committee itself. Enough money has been pledged or donated to run STISP for three years. In Skeet's opinion, however, STISP "should be publicly funded".

STISP has been more than two years in the making, during which time Skeet and his colleagues examined the Office of Technology Assessment in Washington, which advises the US Congress on scientific matter. STISP will be much more modest than its US counterpart. It will be run from a small office by a director, a chief executive and a small secretariat, working to an annual budget of about £30,000 to £50,000.

Henry Gee

Cambridge to host centre

London

BRITAIN'S Science and Engineering Research Council (SERC) has started the process of concentrating scientific research in interdisciplinary university research centres, secure in the knowledge that, despite reservations from sections of the academic community, the scheme has government approval. The site of the first centre, for research into high-temperature superconductivity (concentrating on small devices), announced last week, will be the Cavendish Laboratory at the University of Cambridge. Within the next four months, SERC wants to select sites for three more centres, to be operating by October.

The government is at present considering advice that academic science should be concentrated in a limited number of institutions; against this background, the process of bidding to host a research centre assumes considerable significance. How representative the first centre will be remains a matter for conjecture, given the unpredictable and fast-moving superconductor field.

Cambridge was one of 11 universities invited to bid to host the first centre. Its prize is a six-year grant of £5.3 million, with up to £2.7 million in the first two years, before an initial review. A second, review will be carried out after four years.

SERC expects to support the centre for 6–10 years. The joint SERC/DTI (Department of Trade and Industry) selection committee that considered the bids was said to have been influenced by the efforts Cambridge had made to achieve interdisciplinary collaboration.

The centre will have a permanent staff of four or five researchers, who will be employed by SERC, and a director yet to be chosen) who, will be an employee of the university. The bulk of the research will be done by staff seconded from the host university and elsewhere.

There will be five project coordinators, from the five departments that have already been collaborating on superconductor work at the university — chemistry, physics, engineering, Earth sciences and materials science and metallurgy. SERC stresses that funding for superconductor research will still be available to groups outside the research centre, for which £2 million has been committed in the current academic session.

Bids have been invited for centres in seven other topics, three of which will be selected in the current round of SERC grants. The topics include power applications of high-temperature superconductivity; surface science; molecular sciences; and process stimulation, integration and control.

Simon Hadlington

Vienna to gain biology centre

Vienna

THE University of Vienna announced last week its intention to build a *Biozentrum* in the city's third district. The centre will house five university institutes from the medical and science faculties.

The creation of the centre fulfils an informal promise that the Austrian government made to Genentech and Boehringer Ingelheim when those two companies decided to locate their Institute for Molecular Pathology (IMP) in Vienna. The two institutions will be adjacent and will share support facilities.

The Austrian government will pay about 300 million schillings (\$25 million) for the *Biozentrum*, which will be ready for occupation in 1991. The project is an indication of the changing attitude of the government towards basic research under new Science and Education Minister Hans Tuppy. Tuppy is "the best thing that has happened to Austrian science", said Max Birnstiel, director of the IMP.

The university will endow two new chairs in general biochemistry and molecular genetics. Both positions should be filled in 1988, said Helmut Ruis of the general biochemistry department, with the professors working in existing university labs or at the IMP until the *Biozentrum* is completed.

Some junior faculty members have already been hired, including Rudolf Schwein in yeast genetics and Erwin Wagner in mouse developmental genetics. There will be five university institutes from both the medical and science faculties in the new centre.

The *Biozentrum*, "inspired by" the institution of the same name in Basel, Switzerland, should attract Austrian researchers thinking of returning from abroad as well as foreigners, said Ruis. "It has traditionally been hard to attract people with an international reputation to the University of Vienna", he said.

The IMP will make a difference. Director Birnstiel, who came from the University of Zurich to take the helm in 1986 (see *Nature* 321, 716; 1986), reported that the IMP has almost completed its staff and will move into its new quarters in January, 1988. The official opening of the IMP coincides with a meeting on "Genes in Control of Growth, Differentiation and Disease" to be held at the end of May 1988. The meeting may encourage other good researchers to consider moving to Vienna. But, Ruis warned, the situation may get "tricky" if the quality of research outstrips the money available for grants. It remains to be seen whether the infusion of new life in molecular biology is too little, too late.

Steven Dickman

Late Antarctic spring might be caused by ozone depletion

Washington

The Antarctic spring has arrived weeks later than at any time in the past thirty years, raising fears that the 'ozone hole' has begun to affect world climate.

Radiosonde data relayed directly from the British Halley Bay base reveals that the spring breakdown of the polar vortex — a circulating mass of very cold air that forms over the Antarctic during the long, dark winter — has been "very much delayed", according to Jonathan Shanklin of the British Antarctic Survey.

Normally, the vortex breaks up in late October or early November but this year it has remained until mid-December. Shanklin, one of the first researchers to detect the reduction in Antarctic ozone, suggests that there may be a direct relation between the late spring and the ozone hole. Ozone absorbs ultraviolet light and helps to heat up the atmosphere and break down the polar vortex; with so much less ozone the atmosphere warms more slowly.

The effect will tend to perpetuate itself.

Global CO₂ action

Washington

IN the second year of trying, Senator Joe Biden (Democrat, Delaware) and others have succeeded in attaching an amendment to the current US State Department authorization bill that requires the Secretary of State to work towards concerted international action on the build-up of greenhouse gases in the Earth's atmosphere. The State Department must work with the Environmental Protection Agency to assess the concentrations of such gases, especially carbon dioxide, and the danger they pose, and to begin talks with other countries with the aim of establishing a global policy, perhaps modelled on the recent Montreal protocol for ozone-destroying pollutants.

Whereas chlorofluorocarbons have caused the Antarctic ozone hole, and acid rain has damaged forests, carbon dioxide has not yet been proved guilty of any demonstrable ills. Gathering a political force against it may thus be harder. Moreover, carbon dioxide is released as a direct consequence of combustion, not as a side-effect, and reduction of emission demands policies to conserve energy and to develop alternative sources of power, both of which programmes have lost momentum now that oil is cheap again. But concern over global atmospheric change has been in the public mind lately, and supporters of Biden's amendment think the time is right for the United States to wield its influence.

David Lindley

With a cold atmosphere, stratospheric cloud will be more likely to form and provide the right conditions for chemical reactions that further reduce ozone levels.

The immediate effect of the late breakup of the polar vortex will be limited to Antarctica. But the change is a major one in atmospheric circulation, and "it is likely that it will have consequential effects in the Southern Hemisphere generally", says Shanklin. Modellers will now be busy seeing if there are detectable effects on surface temperature distinguishable from normal variability.

The delayed spring breakup of the ozone hole means that Antarctic organisms will have spent longer exposed to higher-than-normal levels of ultraviolet light. What effect that might have on plankton and Antarctic food chains is unknown.

Alun Anderson

Leningrad tie with California

San Francisco

OFFICIALS from Leningrad University last week visited the University of California (UC) to sign a three-year agreement for academic exchange.

Leningrad University, the Soviet Union's second-best university after Moscow, is known for its physical sciences and its energetic efforts to act as host to foreign scholars (see *Nature* 329, 796; 1987).

The programme, scheduled to begin in the summer or autumn of 1988, provides for the exchange of up to 100 students, professors and researchers each year in the fields of mathematics, physics, language, sociology and education. The exchanges will range from 3-week summer cultural visits for undergraduates to 30-month exchanges for researchers.

Both universities have active foreign exchange programmes. Leningrad University hosts hundreds of scholars annually, from 27 universities, mostly in Eastern bloc countries. Each year about 1,200 UC students study abroad, at 70 host institutions in 28 countries, and UC hosts about 300 foreign students annually at its campuses.

A UC spokesman said the impetus for the programme arose from the cultural accord signed during the 1985 US-Soviet summit in Geneva, which called for increased "contacts, exchanges and cooperation". The programme is the second between Soviet and US universities — the State University of New York has been conducting academic exchanges with Moscow University since 1974.

Marcia Barinaga

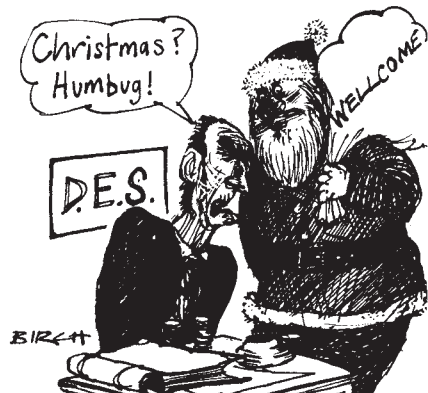
Better pay for UK students?

London

TRADITIONAL ways of postgraduate student support in Britain have been undermined by a new scheme launched by the Wellcome Trust, a private charity. The scheme awards students a much more generous stipend than that offered by the state-funded research councils.

The Wellcome Trust, whose funds for the support of basic medical research were substantially increased last year by the sale of a third of its shareholding in the Wellcome Foundation, the drug company, has begun supporting graduate students in the belief that support by the research councils is now inadequate. The trust plans to spend £1 million on 30 studentships over the next three years.

The value of a Wellcome stipend will be £4,000, with additional funds of up to £2,000 for expenses. This puts the



research councils' basic maintenance grant of just over £2,700 in a very poor light. The Trust says that the value of the research councils' stipend is now so low that promising students are deterred from considering an academic career. The Medical Research Council (MRC) now has a chronic shortage of graduate students (see *Nature* 330, 595; 1987).

Far from being an embarrassment, MRC says that the initiative is welcome. The research councils are now considering whether to raise the level of support to meet Wellcome's challenge. To do this, the MRC would need to inject £2.2 million per year into its postgraduate support programme to maintain student numbers at present levels. If it does not follow suit, the coexistence of research council and Wellcome students would inspire "a degree of tension" in laboratories, says Barton Dodd of the MRC.

Even though the Wellcome scheme will cream off about 15 per cent of the annual graduate intake into the biomedical sciences, neither Wellcome nor MRC sees the scheme as a threat to the dominance of the research councils in postgraduate support.

Henry Gee

British ocean research looks to its long-term future

London

THE British Natural Environment Research Council (NERC) has set as one of its objectives "to predict the oceans", according to Dr John Woods, director of marine sciences at the research council, describing last week NERC's plans for marine research.

At the same time, NERC which is the chief source of funds for geophysical research in Britain seems to be aware that it may be "taking a risk" in advertising such a long-term goal when the government is asking for the solution of practical problems and when it is itself caught in a financial crisis (see *Nature* 329, 96; 1987).

NERC says it is determined to convince the authorities of the need for long-term research, but support also comes from last year's House of Lords report on marine science, which called for closer coordination of government spending. NERC has agreed to provide the secretariat for the new Coordinating Committee for Marine Science and Technology, the centrepiece of the government's response to the House of Lords.

Woods likens oceanography to meteor-

ology, saying that oceanographers need computer models of oceanic circulation comparable with those now used by weather forecasters. He believes there will be sufficient computer power in twenty years to support the more exacting demands of oceanography.

NERC is therefore supporting the development of high-resolution oceanographic models and the technology of data-gathering using unmanned deep ocean vehicles as well as remote sensing from satellites and acoustic sensors.

Given financial uncertainties, NERC is planning a core programme in marine science and a menu of options to be carried out if funds are available. In the core are a plan to model the North Sea the better to predict the transport of pollutants and a project to construct a fine-resolution model of the Southern Ocean.

A project to develop two types of unmanned submersibles (under the name of AUTOSUB) should yield data of a novel character from the deep oceans, but NERC is also committed to studies of the geochemistry and global circulation of the oceans.

Philippa Lloyd

Troublesome trials for AIDS vaccines

Washington

THE first US clinical trial of a vaccine against AIDS (acquired immune deficiency syndrome) is having difficulty finding enough suitable volunteers. The trial — administered by the US National Institutes of Health (NIH) using recombinant gp160 protein produced by the biotechnology company MicroGeneSys — began in October. But so far, only 26 volunteers out of the required 81 have been found for the programme, raising fears for the future of AIDS vaccine testing.

Frank Volvovitz, president of MicroGeneSys, says that hundreds of male homosexuals came forward for the trial, but most were rejected for medical reasons such as hypertension. Others dropped out because of the time commitment required.

Volvovitz says the dearth of participants in the trial is partly attributable to volunteers' reluctance to become seropositive for HIV (human immunodeficiency virus), the virus that causes AIDS. Although blood from volunteers will test positive for HIV after vaccination, a Western blot test — routinely performed to screen out false positives — would prove that the person had not actually been infected with the virus. Further, the NIH will issue certificates to those who participate in the trial, to prevent discrimination against them on the basis of seropositivity. The present trial is

meant to test only safety and immunogenicity. Later, much larger groups will be needed to assess the vaccine's efficacy.

Gerald Quinnan, head of virology at the US Food and Drug Administration, said that an "endpoint" other than the diagnosis of AIDS should be found to indicate the failure of a vaccine. The long incubation period of AIDS means waiting for years to see if subjects are protected from inadvertent exposure to the virus.

Quinnan also pointed out that the number of homosexual males in the United States will not be large enough to support the widespread testing of multiple vaccines, and new study populations will need to be investigated.

As the availability of study populations tightens, developed countries may be tempted to look at poorer countries as places for testing vaccines. The World Health Organisation (WHO) has formed a committee on vaccines as part of its new AIDS programme to help less-developed countries decide among the many proposals for testing AIDS vaccines with which they are likely to be presented.

Sir James Gowans, chairman of the committee, says the WHO should be able to provide an independent assessment, and although the body has no regulatory power it would use "moral force" to prevent exploitation.

Carol Ezzell

Boycott of South African science

SIR—We are all white scientists and, like Festenstein, Sachs, Hepple and Shall (*Nature* 328, 570; 1987), three of us were also born and bred in South Africa. However, despite the fact that we abhor apartheid, we choose to live and work in this country. In essence, Festenstein *et al.* argue that if one sincerely opposes the system, there are only two moral options open to whites in South Africa: imprisonment or emigration. Willingly to choose imprisonment requires a degree of courage we do not profess to have, but, on the other hand, emigration can be (and often is) construed as a purely self-interested act by black and white alike.

Although we acknowledge that the most fundamental need is to rid this country of apartheid, we feel that we cannot contribute to this cause by leaving and so distancing ourselves from a problem inherited from history. In this connection, we note that the ANC (African National Congress) and the UDF (United Democratic Front) have urged white South Africans to *stay* and contribute to the elimination of apartheid. Assuming that the majority of people who have left this country since Sharpeville (1961) did so for moral reasons — and not fear for their security or of unpalatable change — then certainly these progressive-minded people could have had only a positive effect on the future of this country if they had remained here. It is one thing to urge the international scientific community, from abroad, to shun “the self-interest of those who enjoy the privileges of white minority science” and quite another to try to face up to the problems of a country in conflict by physical presence.

We work at an institution that unconditionally rejects racial segregation and is committed to nondiscrimination in the selection of both its students and staff, as well as going to some (enough?) lengths to accommodate the less privileged, despite government censure. We accept the fact that in South Africa, neither the individuals nor the academic institutions they represent can live in isolation from the dominant social and political trends in our society; that by working in this country one can distance oneself from apartheid only by attempting to oppose it. But we believe that research and teaching which are not directed towards support of apartheid and, where possible, attack the roots of the system (prejudice, ignorance and economic inequality), have an important role both now and in the future. Surely continued cultural and scientific endeavour will benefit a new South Africa. We believe that part of our aim as scientists in this country should be to help with the advancement and education of the less privileged South Africans. We

therefore reject statements questioning our integrity by those who have chosen the option of emigration.

The major issue is that developed countries (and those South Africans now living in them) should not seek to soothe their collective sense of outrage by attempting to destroy legitimate scientific endeavour and teaching in South Africa. That would be a cruel kindness indeed and one from which it might take many generations to recover, after the last vestiges of current discrimination have been removed. Would Festenstein *et al.* then be willing to return to help rebuild this shattered land?

In reply to the more general debate about the boycott of South African science (J. Maddox *Nature* 327, 269–276; 1987; J.G. Wilson *Nature* 328, 288; 1987; W.D. Stein *Nature* 328, 374; 1987) we would like to raise the following points:

(1) It may be naive to think that a scientific boycott would have the ‘desired’ effect of bringing South African science rapidly to its knees. Rather, we believe that, as has been the case with the sports, cultural and the partial technical boycotts, South African science under pressure of a boycott will ‘hang in there’ for years to come. Despite what Wilson may think, the boycott will not be the agent of extinction for the ‘dinosaurs’, but rather one of regressive evolution. (2) The rapid urbanization — that is, the downfall of the apartheid ideology — has been brought about by the search for jobs provided by science and technology. We therefore believe, unlike Stein, that a total standstill of South African science for 5–10 years would have a detrimental effect on the living standards of the underprivileged majority of our population.

Frontier technologies, especially in the minerals extraction industry, are the main breadwinners for South Africans. Many of these workers and their dependants need scientific and technical training fully to realize their potential so the current demand in South Africa is therefore for more and better education, not for a destruction of education in general. Additionally, South Africa is, particularly in the rural areas, a developing country in dire need of appropriate development in science and technology, particularly in the fields of water, agricultural and other natural resource management. A halt of advancement in these fields will hurt everyone, but least of all the white communities, whom Stein assumes will be hurt almost exclusively. Wilson has clearly grasped this point, but is wrong in assuming that we necessarily belong to a group of scientists who would secretly welcome an across-the-board scientific boycott.

However, the debate about a potential academic boycott has had very positive

effects within the scientific community. Stein has understood this remarkably well. The very idea of such potential actions has made many scientists, (ourselves included) aware of their vulnerability. Through the internal actions of Academic Staff Associations, at the more liberal universities some of these scientists may soon be forced openly to take a stance on the issues that confront us. Such an enforcement may possibly open the door for a type of selective boycott such as an academic ‘Sullivan Code’, as visualized by Stein (however, we feel that such a code would have to be applicable to educational and academic rights violations worldwide). Rather than across-the-board boycotts, we ask for the help of our colleagues abroad in establishing more positive ways of eliminating inequality, for example in creating affirmative action programmes at our academic institutions.

This country needs to continue to increase its efforts over the next few decades towards training black scientists and technologists — many of whom will be taking up executive positions in internationally competitive concerns — if it is to develop a reasonable living standard for all its people. An across-the-board scientific boycott would seriously hinder this goal, leading to a massive and irreversible ‘brain-drain’ (which is already occurring) and setting this country on a further slide towards greater entropy. At that point it will definitely be at the mercy of foreign economic exploiters and human-rights violators for an untold length of time. Apparent political freedom will then be hollow indeed.

Finally, we have discussed the academic boycott issue and the contents of this letter with several of our black colleagues. The absence of their signatures reflects disagreement with some (but not all) the points we have made, as well as a fundamental difference of opinion as to the value of debating academic boycotts. Nevertheless, one positive result of this correspondence is that it has led to an expansion of dialogue with our black colleagues on some of these very fundamental issues.

MAARTEN DE WIT

JOHAN KRUGER

Bernard Price Institute of Geophysical Research,

RODGER HART

MARIAN TREDoux

NICK LINDSAY

Schönland Centre for the Nuclear Sciences,

*School of Earth Sciences,
University of the Witwatersrand,
Witwatersrand 2050,
Johannesburg,
South Africa*

• This correspondence is now closed.—
Editor, *Nature*.

Where science has gone wrong

SIR—Only the most casual reading of Popper's writings can have misled Theocharis and Psimopoulos (*Nature* 329, 395; 1987) to summarize Popper's epistemology as requiring "the Earth is a flat disk" to be a valid scientific statement in contrast with "the Earth is (approximately) a sphere". Statements concerning singular empirical facts, such as the latter, are the rocks and foundations of Popper's epistemology, but a simple list of true empirical facts does not constitute a scientific theory.

Scientific models, hypotheses or theories try to condense statements relating to subsets of such sets of empirical facts into fewer, logically consistent statements, which have to reproduce by logical deduction at least their generating subset of facts (In the latter case, this set of statements should be considered only a valid model.) The fewer the statements needed in order to formulate a theory and the larger the subset of "explained" facts, the more potent is the theory. As lists of empirical facts are potentially infinitely large, theories phrased as negations are usually the most powerful. ("There are none but spherical celestial bodies" is such a scientific, albeit invalid, theory.)

Popper's demand of falsifiability requires a *theory* (as opposed to a statement of a singular empirical fact) to have among its logical deductions at least one statement concerning a new, single, empirical fact (for example, "The Earth is a flat disk"), which in principle can be contradicted by a statement about an empirical observation or the result of an experiment. A case in point is the report by Greenough and Harvey (in the issue of *Nature* in which Theocharis and Psimopoulos appear, p.585) concerning the refutation of the "neutral theory of evolution" by a cleverly designed experiment.

Unless this theory can be modified to incorporate this new finding, it should be considered a falsified (but still scientific) theory. This example also nicely demonstrates Popper's claim that (unless I run my laboratory by throwing dice) every experiment is, at least to some degree, 'infected' by an underlying theory, which the experiment tries to corroborate or, preferentially, to refute. Without the "neutral theory of evolution", the experiment to which I refer would never have been made.

In sharing the author's evaluation of the other "philosophies" discussed by them and their concern about the decline of the public standing of science, I would suggest as the only remedy that scientists adhere much more closely to the principles of Popper's epistemology. The late Sir Peter Medawar, whom the authors quote

approvingly, held Popper's epistemology in the highest esteem and in all probability would have agreed with any advice. For publicly calling Popper a "betrayer of the truth" the authors (and *Nature*?) owe him a public apology.

RAINER FACIUS

DFVLR, Institute of Aerospace Medicine,
Biophysical Branch,
Linder Höhe, 5000 Köln 90, FRG

SIR—Theocharis and Psimopoulos state (allegedly paraphrasing Lakatos), that "if observations are theory-laden, this means that observations are simply theories". Rubbish.

If, whenever I combine my visual observations with ballistics theory *A*, the falling cricket ball hits me on the head; but whenever I use ballistics theory *B*, I catch the ball; then I shall not yet consult an optician and will (provisionally) retain theory *B*. Conjectural it remains, but for me at least, it is superior to the apparently refuted theory *A*. And the bruises on my head enable me to distinguish between theories and observations — or, in this example, between the ballistics and the balls.

Thus reasons Popper; thus indeed does science advance by way of an always incomplete search for greater predictive reliability; and thus is scientific method defined.

To lump together Popper and Feyerabend is perverse; they are as unlike as Kant and Hegel. The attack on Feyerabend's anarchic delusions I could accept, and perhaps he has contributed to that value-eroding "relativism" which Allan Bloom has recently and elegantly castigated¹. But Popper's three great early works²⁻⁴ — which Theocharis and Psimopoulos do not cite — remain for many the most scholarly and rigorous defence and definition, not only of scientific method, but of democracy. They are products, moreover, of an Austrian emigré who knew too well the potential damage which philosophical folly and self-confident certainty can wreak on a society, as surely as Lysenko could (with Stalin's backing) blight Soviet science.

MARK F. CANTLEY

131 rue Verbist,
B-1030 Brussels, Belgium

1. Bloom, A. *The Closing of the American Mind* (Simon & Schuster, New York, 1987).
2. Popper, K.R. *The Open Society and its Enemies* (Routledge & Kegan Paul, London, 1945).
3. Popper, K.R. *The Poverty of Historicism* (Routledge & Kegan Paul, London, 1957).
4. Popper, K.R. *The Logic of Scientific Discovery* (Hutchinson, London, 1959).

SIR—Theocharis and Psimopolous have described the damage done to science by irresponsible philosophical scepticism and its popular derivatives. A similar point could be made about the destructive effects on personal and political morality. History shows the murder-

ous consequences of the moral equivalents of the epistemological 'isms' that Theocharis and Psimopolous criticize; and of course scientists are as easily seduced by anyone. In science, as Theocharis and Psimopolous suggest, and also in morality, the frivolous sceptic can often be caught in self-contradiction. Philosophical complacency, however, will not do; contrary to what both sceptics and conservatives often seem to believe, philosophical questions do matter.

Surely, though, Theocharis and Psimopolous are unfair to Popper in the example they use to illustrate his falsifiability criterion. The statements "the Earth is (approximately) a sphere" and "the Earth is a flat disc" are, in principle, both falsifiable (if 'approximately' is defined). They differ only in that the second has been amply falsified, while the first has not been, and (it is impossible to avoid begging the question) is never likely to be.

GRAHAM J. KEMP

Department of Human Metabolism
and Clinical Biochemistry,
Medical School, Beech Hill Rd,
Sheffield S10 2RX, UK

SIR—I was a contributor to the BBC's *Horizon* programme "Science . . . Fiction", of which Theocharis and Psimopoulos complain. Their article attacks the recent philosophy of science and, by implication, sociology and history of science as well. It contains many mistakes and *non sequiturs*.

Quoting Mrs Shirley Williams, the authors blame diminution of research funds on the failure of science to deliver increased wealth. There may be some truth in that. Since the Second World War, the most lavish spending on pure science has been on fundamental physics and astronomy, which have not delivered much commercial value. But it is odd to blame that failure on physicists' and astronomers' poor grasp of scientific method; there has not been a great deal wrong with the physics and the astronomy, so far as I know.

Likewise, can it be that the excessive funds expended on science in the service of the military arise out of the better grasp of epistemology at the Royal Radar Establishment at Malvern than at Jodrell Bank? The argument that the failure of the sciences to win the support they claim is due to scientists' poor understanding of epistemology is plainly silly.

The authors' assertion that scepticism "entails social, political and every other kind of anarchism and disorder" is dangerously naive. Sceptical ideas can be used to justify wholesale change, but also moderation. Thus Karl Popper attacks totalitarianism because there is no possibility of establishing a science of society sufficiently reliable to justify the moral costs of large-scale social engineering. The opposite side of the coin is that pro-

gressive models of science can be used to justify democratic institutions, or abhorrent political ideas, such as the "scientifically" inspired and executed "final solution" to the "scientific" problems of racial impurity.

Their "third danger", that sceptical approaches stifle progress, may have a grain of truth, and scientists would be ill-advised to engage in too much philosophical sophistry at the laboratory bench. Nevertheless, the possibility of good work is not destroyed by reflection upon the complex nature of science. It is a mistake to think that philosophy can contribute directly to scientific methodology. It is an equally grave mistake to think that the naive motivating ideology of science can provide epistemological foundations. Such aspirations died with logical positivism.

The only thing that makes clear good sense in Theodorich and Psimopoulos is the claim that the privileged image of science has been diminished by the philosophical, historical and sociological work of past decades. One hopes this is the case. Grasping for special privilege above and beyond the world we make for ourselves — the new fundamentalism that Theodorich and Psimopoulos press upon us — indicates bankruptcy of spirit luckily not yet widespread in the scientific community.

HARRY COLLINS

Science Studies Centre,
University of Bath,
Claverton Down, Bath BA2 7AY, UK

AIDS: Incubation or latency?

SIR—It is commonly assumed that the clinical manifestations of AIDS (acquired immune deficiency syndrome) follow an 'incubation period' which may be as short as months or as long as many years. 'Incubation' indicates a continuous process that, starting with its determinant cause, will produce its full effect at a relatively fixed time in the future, like the incubation of an egg or of some infectious diseases. To speak of incubation, one must be reasonably certain that the process has a continuity in time that irrevocably links the effect to its cause. When the final effect requires, besides the initial cause, some further condition that may or may not arise in the future, we speak of 'latency'. The manifestation of a disease later in life following the presence of an inherited predisposing characteristic is typical of this case. The exact course of the development of AIDS following contact with the virus is not yet known, but the long and highly variable time between infection and symptomatic course clearly indicates that this is a case of latency rather than incubation.

The distinction of latency from incubation is not simply pedantic. The prolonged absence of symptoms in a large majority of those that have had contact with the virus is already indicative that in many — probably most — of them the disease will never develop. The determination of the causes that end the latency of AIDS and bring about its symptomatic course can only be accomplished by a continuing investigation of the cases in which the disease is latent.

It therefore requires the full cooperation of those who are immunologically positive but have no clinical manifestations. Although the public health authorities of many countries, the United States first and foremost, have employed every means to inform the public of the seriousness of the disease, they have done nothing to convince it of the interest and importance of the participation in the research and prevention of the disease of those who are immunologically positive but have no manifest symptoms. In fact, the prejudices of the public against those it sees as potential carriers of the disease have been allowed to reach a point where a person who knows himself to be immunologically positive and has no other symptoms of AIDS would be well advised to hide the fact.

Proper information of the public, and most important the need to regain the confidence of those who have had contact with the virus but show no clinical symptoms, require that the implications of the AIDS infection be honestly presented. A suitable beginning would be the acknowledgement that infection is followed by an indetermined period of latency and that in most cases this may well turn out to be as long as the natural life of the individual.

GREGORIO WEBER

Department of Biochemistry,
University of Illinois,
Urbana, Illinois 61801, USA

Leprosy vaccine

SIR—I wish to comment on your news article on a new vaccine trial against leprosy being undertaken in Maharashtra, India¹. Recognizing that a third of all patients with leprosy are in the Indian sub-continent, the government of India has made a major commitment to the control of this disease. The UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases has, through its immunology of leprosy (IMMLEP) programme, supported a number of research efforts to develop candidate vaccines against leprosy. A mixture of killed *Mycobacterium leprae* together with live BCG vaccine has been shown to be effective in preventing infection in animals. It is currently being tested in controlled trials in 29,000 contacts of leprosy patients in Venezuela and in a

general population of 120,000 in the northern region of Malawi.

I wish, however, to correct two mis-statements in your article, concerning the possibility that this vaccine may contain retroviruses and foreign DNA. In the light of the stringent procedures for purifying the bacilli from armadillo tissue, including their exposure to DNAase and extensive washing, followed by gamma radiation (2.5 Mrad) and autoclaving^{2,3}, the statement that this vaccine contains retroviruses and foreign DNA is entirely incorrect and unfounded.

R.J.W. REES

WHO IMMSEP M. Leprae Bank,
National Institute for Medical Research,
London NW7, UK

1. Jayaraman, K.S. *Nature* **328**, 660 (1987).
2. WHO Report of the fifth meeting of the Scientific Working Group on the Immunology of leprosy. Protocol 1/79. TDR/SWG/IMMLEP (5) 80.3 (1980).
3. Rees, R.J.W. *Int. J. Leprosy* **51**, 515–518 (1983).

Genome mapping

SIR—It is more in sorrow than in anger that I protest at the headline of your news story about our genetic linkage map ("Critics denounce first genome map as premature", *Nature* **329**, 571; 1987). The headline seems strangely disconnected from the facts even as they are detailed in your account.

After correctly noting that we will make our probes freely available for research purposes, you report, "[Collaborative] is, however, applying for patents on the probes and stands to gain if (there are) . . . commercial applications".

Let me assure you that this entrepreneur fervently hopes we stand to gain. That is precisely why we invested \$12 million in this enterprise. *Nature* goes on to talk of "cries of commercialism", of anonymous stock analysts who criticize us for an approach to genome mapping "too grand and expensive" and of unnamed critics who accuse us of "scientific merchandising" for reporting our accomplishments.

We confess that we do want to make a buck, honestly. We think investing our shareholders' money in genetic linkage mapping is a good idea for science, and a good idea for our business. I think too much is being made of the gaps and not enough of the map.

If Collaborative Research makes money doing all this, colour us happy and colour those stock analysts green.

ORRIE M. FRIEDMAN

Collaborative Research Incorporated,
Two Oak Park,
Bedford,
Massachusetts 01730, USA

Letters submitted for Correspondence should be typed, double-spaced, on one side of the paper only □

What the solar nebula was like

Two remarkable papers in this issue are a vivid pointer to the ways in which it may be possible to learn how the material from which the Solar System formed was itself produced.

WHAT is known about the solar nebula, the cloud of gas and dust from which the Solar System formed? And where did that come from? The two papers on pages 728 and 730 of this issue from Dr Edward Anders and his colleagues at the University of Chicago and Washington University, St Louis, are an intriguing pointer to how those legitimate questions may eventually be answered, but are also an illustration of how complicated the questions may prove to be.

The new development, ostensibly that of the discovery of silicon carbide in the Murray meteorite, is not in itself surprising. With the recognition that small crystals of diamond are to be found in meteorites, there are few who will be surprised to learn that there is silicon carbide as well.

In any case, as Anders and his colleagues say, silicon carbide has been recognised spectroscopically in the atmospheres of stars, so that the discoveries now reported are the first direct demonstration that the spectroscopic identifications are correct (as well as a vivid sign that the minerals carried by meteorites do indeed derive from stars). But it now also emerges that the occurrence of this material has a direct bearing on the chemical composition of the solar nebula or, more accurately and significantly, on the nature of at least one of the sources from which it was formed.

None of this should diminish wonder at the technical feat now reported, which is a breathtaking illustration of what microanalysis can now accomplish (see cover of this issue). Those concerned have been working with small grains, some just a few hundred Ångströms in diameter, which have themselves been extracted from a sample of the Murray meteorite after treatment with strong acids and oxidizing agents to remove all but the most refractory of structures.

Even so, by the evidence of the two papers, it has been possible to identify some of these grains as sources of neon and xenon with anomalous isotope composition, a superbly delicate exercise in mass spectrometry. Telling that some of the grains consist of silicon carbide (most directly by Raman spectroscopy, most convincingly by electron diffraction) may be, by comparison, almost child's play. Even so, it is striking, and significant, that some of the crystals are found to be twinned, which fits in well with what

emerges from the sequel, the suggestion that the silicon carbide grains were formed as such in the atmosphere of some star.

But what kind of star? This is the sense in which the mere presence of silicon carbide crystals is significant, by the test of what the chemistry textbooks call the principle of mass action. Atoms of silicon in a mixture of carbon and oxygen will preferentially be linked to oxygen rather than to carbon. For silicon carbide rather than silica crystals to be formed by condensation in the atmosphere of a star, carbon must be abundant relative to oxygen, which Anders and his colleagues say implies a star rich in carbon, presumably in the late stage of its evolution.

That is intriguing because it conflicts directly with the ratio of carbon to oxygen in the Solar System as it is, and where oxygen is half as abundant again as carbon. The implication is that the star or stars in whose atmosphere the grains of silicon carbide were formed were only minor contributors to the solar nebula. That, in itself, is no great surprise. The visible regions elsewhere in the Galaxy at which stars appear to be being formed are places where there is much more material than required for a single star, with the result that groups of stars formed at the same time. Why should not the Solar System also have been formed from a nebula to which several stars had independently contributed?

These are merely the straightforward implications of the presence of silicon carbide in the Murray meteorite. Anders and his colleagues are after much more specific information about the character of the stars that may have contributed material to the solar nebula, and which they hope to win from the evidence gathered by their microprobe analysis of the isotope composition of the silicon carbide. (Mass spectrometrists, used as they are to the manipulation of small samples, may well pause to note that Zinner, Tang and Anders laconically explain, on page 730, that their microprobe measurements have entailed the analysis of atoms thrown off a μm -sized patch on the surface of an aggregate with sensitivity good enough to discriminate between $^{13}\text{C}^{14}\text{N}^-$ and $^{12}\text{C}^{15}\text{N}^-$.)

It is important for the argument as so far advanced that the Murray meteorite has yielded two kinds of grains containing silicon carbide, in one of which it is associated with an amorphous compound

of silicon and oxygen that may (this is surmise only) originally have been silicon nitride. That would have been oxidized by the hyperchromate treatment used to remove graphite and other materials containing organic carbon.

The most striking feature of the data now reported is that the isotope anomalies discovered by the microprobe analysis are, by the standards of terrestrial isotope analysis, huge — some samples yield an excess of ^{13}C measured by $\delta = +6,000\%$.

The conclusions should provide the astrophysicists who study nucleosynthesis with some tantalizing puzzles with which to grapple over the holidays ahead. First, there seem to be at least two sources of anomalously heavy carbon represented in the meteorite sample — one of them enriched in ^{13}C six times more than the other, but each of them associated with nitrogen deficient in the heavier isotope ^{15}N . Which of the anomalous noble gas materials is associated with which kind of anomalous silicon carbide is not unambiguously clear at this stage, although it appears that anomalous xenon (enriched in ^{136}Xe) is linked with the source of less markedly heavy carbon.

The silicon anomalies now reported will be even more taxing for the nucleogenesis community, which will have to account for at least three kinds of sources — producing silicon both enriched and depleted in ^{28}Si as well as silicon enriched in ^{29}Si . Anders and his colleagues explain that the two sets of anomalies are independent of each other — carbon and nitrogen anomalies must say something about the hydrogen-burning phase of a star's evolution and the silicon anomalies something about the later phases. But what?

Time will tell. This is not the first occasion when people have based inferences about the solar nebula on data gathered from materials at the surface of the Earth. The abundance of ^{26}Al has, for example, been regarded for more than a decade as an indicator of short-lived ^{26}Mg in the material of the solar nebula, suggesting a supernova explosion immediately preceding the Solar System.

The wealth of detail likely to emerge from the further research of Anders and his colleagues promises to specify the nature of the stars contributing material to the nebula, while the ^{26}Al evidence will suggest to some the event that swept this material into concentrations from which stars could form.

John Maddox

Oceanic evolution

Rifting or drifting in the Red Sea?

Enrico Bonatti

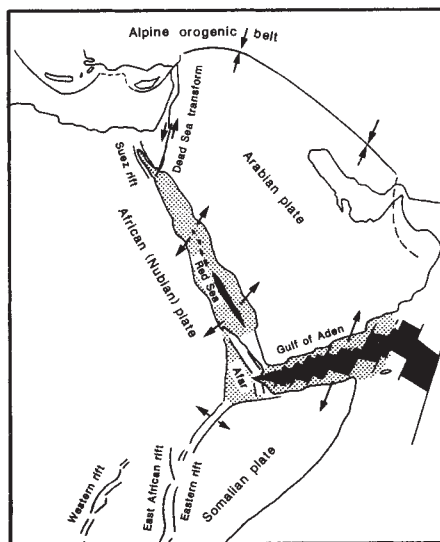
THE Red Sea has inspired not only one of the better-known instances of 'geomythology' (when it parted for Moses), but also one of the first recorded examples of sound geological reasoning: in 400 BC, Herodotus estimated that it would take 10,000–20,000 years for the Red Sea to fill up with sediment if the Nile were diverted into it. Even today it can be difficult to distinguish geological from geomythological thinking where the Red Sea is concerned. The Red Sea, Gulf of Aden and East African rifts are central to the study of plate tectonics, because here we can observe the breakup of a continent, the transition from a continental to an oceanic rift and the initial steps of seafloor spreading. On page 716 of this issue, R. W. Girdler and T. C. Southren discuss the question of whether the floor of the northern Red Sea consists of old, stretched, continental crust, or newly injected oceanic crust.

The different rifts that converge in this region can be regarded as representing different stages in an evolutionary sequence. In the East African rift, rupturing of the continental lithosphere has begun, but very little extension has taken place. The Gulf of Aden, on the other hand, is clearly oceanic: in its eastern part, the continental lithosphere separated about 10 million years (Myr) ago, and the separation has subsequently progressed westwards, with the tip of the oceanic rift now impinging against the African continent (at Afar) in the Gulf of Tadjoura (see figure). The Red Sea rift occupies an intermediate position in this evolutionary scheme. It is generally agreed that its opening has progressed from south to north; however, the extent and mode of its opening remain controversial. It has been proposed that the Red Sea is carpeted by oceanic crust over its entire width, from coast to coast, which implies a long history of seafloor spreading in the Red Sea.

In one formulation, first proposed by R. W. Girdler and P. Styles (*Nature* **247**, 7–11; 1974), the evolution occurs in two stages, the first starting as far back as 40 Myr ago. This view implies also that the boundary between continental and oceanic lithosphere is relatively sharp with a narrow transition zone. In contrast, other groups suggest that seafloor spreading started recently (5 Myr ago) in the southern Red Sea, and that oceanic crust is absent in the northern Red Sea. In this view, outside a narrow oceanic axial zone, the Red Sea is carpeted by stretched and thinned continental crust injected diffusely by basaltic dykes (see figure). Both

views are compatible with plate-tectonic reconstructions of the region (see, for example, Joffe, S. & Garfunkel, *Z. Tectonophysics* **141**, 15–22; 1987).

If we consider the Red Sea to be a passive margin in the making, it is obvious that the contrasting views on its mode of opening have important implications for our ideas on the formation of passive margins in general, such as those at the edges of the Atlantic Ocean. This is of



Simplified, schematic plate-tectonic framework of the Red Sea/Gulf of Aden area (modified from Joffe, S. & Garfunkel, *Z. Tectonophysics* **141**, 15–22; 1987). Black areas, oceanic crust in the Gulf of Aden and Red Sea. Stippled areas, transitional crust, consisting of stretched and thinned continental crust injected by diffuse basaltic dykes. According to a different viewpoint, discussed in the text, most of the area shown here as covered by transitional crust is carpeted by oceanic crust.

more than academic interest, because the thick sediment piles that have accumulated on several passive margins are a main source of petroleum.

Girdler and Southren in this issue discuss the geological evolution of the northern Red Sea, where a crustal extension of more than 100 km is required by well-documented movements along the Gulf of Aqaba/Dead Sea strike-slip fault which started in the early Miocene (about 25 Myr ago). But has this much extension occurred mostly from the creation of new oceanic crust, or from the stretching of the continental crust and diffuse injection of it by basaltic dykes? Girdler and Southren present new evidence for those who favour the first explanation. Their ammunition consists mainly of magnetic and gravity data, together with a review of available seismic-refraction data. The

data are consistent with the hypothesis of a mostly 'oceanic' northern Red Sea; the same data, however, are also consistent with the 'stretching' view. For instance, the lack of high amplitude, Vine-Matthews magnetic anomalies in the northern Red Sea is in line with the absence of oceanic crust there. Alternatively, as Girdler and Southren suggest, the absence of these anomalies could result from very slow (about 0.5 cm yr⁻¹) seafloor spreading combined with burial of the crust under a thick cover of evaporite sediment (a thick Miocene layer of evaporite is almost ubiquitous in the Red Sea basin).

Seismic refraction data are equally open to more than one interpretation. They do show a rapid decrease of crustal thickness moving seawards from the coast lines of the northern Red Sea. However, the inferred crustal compressional-acoustic-wave velocities, generally ranging from 6 to 7 km s⁻¹ in the upper part of the crust and suggesting the presence of basic intrusive rocks, do not necessarily mean oceanic crust is present. Basic intrusives injected diffusely into remnants of thinned and stretched continental crust could also explain the seismic data. The ambiguity of these data is illustrated by the following: large areas of the eastern-desert region of Egypt (clearly part of the African continent) are covered by mafic/ultramafic (ferromagnesian-rich) complexes of Pan African age (500–700 Myr ago); if these terrains were under water and subjected to seismic refraction experiments, seismic velocities close to those of the northern Red Sea would be found. Further difficulties for an 'oceanic' northern Red Sea are provided by recent work of F. Martinez and J. Cochran (*Tectonophysics*, in the press) who find, in sections across the northern Red Sea, a block-faulted crustal structure typical of stretched and thinned continental crust seen at passive margins.

The Red Sea controversy is in part a matter of semantics. During the advanced stages of rifting, basaltic injections into the stretched continental crust occur diffusely over a wide zone, which gradually narrows, at the same time as the crust thins, isotherms in the asthenosphere (the zone of the upper mantle beneath the rigid lithosphere) rise and become focused, and transitional basalt changes to mid-ocean-ridge basalt (MORB). Only when the zone of basalt injection has narrowed to a strip along the rift axis a few kilometres wide, are the last remnants of continental crust swept apart, and Vine-Matthews magnetic anomalies develop. By most definitions, the initial formation of oceanic crust occurs at this point. Even before this stage is reached, however, the crust can be largely basic in composition, though not yet oceanic.

Thus, even if a geologist had followed Moses and had been able to walk directly

on the northern Red Sea crust (surely the evaporites were also made to part to facilitate walking across, because they are notoriously slippery), he would have had difficulty deciding whether he was walking on oceanic crust or on some sort of transitional, pre-oceanic crust. The same difficulty faces the geologist today who works in the Afar rift, at the south-west corner of the Red Sea (see figure). Afar is covered largely by basalt, whose composition, however, is different from the MORB that makes the oceanic crust and carpets the oceanic axial trough of the southern Red Sea. No wonder that northern Red Sea seismic-refraction data are open to more than one interpretation. The question of the nature of the crust

in an embryonic ocean like the Red Sea remains unsolved. Other problems can be tackled in the Red Sea region. Are rifts caused initially by thermal anomalies in the upper mantle (active rifting) or by horizontal plate motions (passive rifting)? Why does the initial emplacement of oceanic crust occur not along a continuous crack but at discrete, equidistant nuclei? Why is volcanism intensely active in some rifts and not in others? By what tectonic processes does the lithosphere extend in rift zones? There is room for much exciting geological research, and also, still, for some good geomorphology. □

Enrico Bonatti is at the Lamont-Doherty Geological Observatory, Palisades, New York 10964, USA.

Muscular dystrophy

The missing link in DMD?

Clarke R. Slater

DUCHENNE muscular dystrophy (DMD) is one of the most devastating of all inherited muscle diseases. First apparent in boys in early childhood, it leads inexorably to death, usually by the age of twenty. Many hypotheses have been proposed for the cause of the muscle wasting; some very plausible but all open to challenge because of the lack of one vital piece of information — the nature of the altered or missing protein. Now, after rapid progress in the molecular biology of the X-chromosome during the past few years, a protein encoded by the DMD gene has been identified and is reported¹ in today's issue of *Cell*. Dubbed 'dystrophin', it is present in normal muscle but is undetectable in muscle from patients with DMD. On page 754 of this issue², Hoffman, Kunkel and their colleagues report further that dystrophin is present within normal muscle fibres in association with the triads, the intracellular junctions where electric excitation of the muscle fibre triggers contraction (see figure).

A key to these advances was the discovery that a region of the X-chromosome in mice is very similar to the DMD gene in humans³. Proteins encoded by fragments of this mouse DMD gene were produced in bacteria and used to raise specific antibodies. These antibodies recognize a protein, dystrophin, in normal skeletal and cardiac muscle of both humans and mice. Dystrophin is present in very small amounts (estimated to be about 0.002 per cent of the total protein) but has a relative molecular mass under dissocia-

ting conditions of about 400,000 (400K), consistent with the size of the 14-kilobase messenger RNA product of the DMD locus. Most significantly, dystrophin is undetectable in skeletal muscle from the two boys with DMD studied.

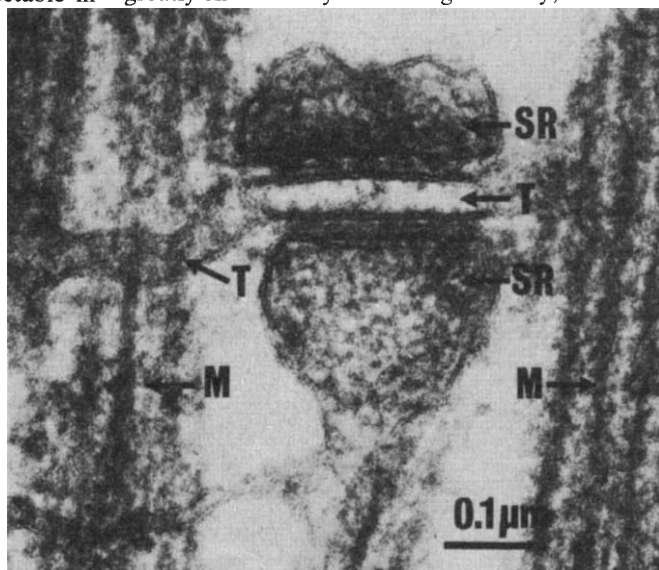
Over the years, several strains of mice have been found with wasting diseases of skeletal muscle. An X-linked form of mouse muscular dystrophy, found several years ago⁴ and named *mdx*, seemed to be a particularly relevant model of DMD, even though affected mice survive an early phase of extensive muscle-fibre death and thereafter have a nearly normal life⁵. The importance of these *mdx* mice is now greatly enhanced by the finding that they,

too, seem to lack dystrophin in their heart and skeletal muscle. Although this finding lends support to the view that the absence of dystrophin is involved in bringing about muscle-fibre destruction, it raises an important question about the devastating effects of DMD: why should the absence of dystrophin cause a fatal disease in boys, but leave mice almost normal?

Studies of the primary structure of dystrophin, predicted from the base sequence of the DMD gene, give some hints about its nature. For one thing, both its relative molecular mass and very low abundance make it clear that dystrophin is not nebulin, recently proposed⁶ as a candidate for the DMD gene product. Nebulin is too large, having an apparent molecular mass of about 500K. Moreover, although it is abundant in normal human muscle (about 3 per cent of total protein) and apparently much less so in DMD muscle, it appears to be present in normal amounts in the muscles of *mdx* mice¹.

More positively, the strong similarity between the dystrophin genes of mouse and man implies that its detailed structure is vital for its function. In each species, the predicted amino-acid sequence suggests that a substantial part of the amino-terminal region has an alpha-helical structure³. This could be consistent with an affinity either for cell membranes or for helical proteins such as myosin. More recently, it has been found that there is significant similarity between the amino termini of dystrophin and of alpha-actinin, a protein that seems to bind actin filaments to cell membranes⁷. Perhaps, as the authors of the paper in this issue² suggest, dystrophin is a link in a more than metaphorical sense, binding the contractile filaments to the internal membrane system of the cell, and thus linking the membranes from which calcium is released on muscle excitation to the contractile proteins that are activated by it.

To learn where dystrophin resides in normal muscle fibres, Hoffman *et al.*² used their antibodies to probe subcellular fractions of normal mouse and rabbit skeletal muscle. Their results suggest that whereas dystrophin is not associated with insoluble filamentous proteins, it is associated with intracellular membranes and particularly with the triads. Hoffman *et al.* identified fractions enriched for triads using probes for two of their key membrane proteins: the 'voltage sensors' of the transverse tubular system and the calcium pores of the sarcoplasmic reticulum. During subcellular fractionation, dystrophin, although clearly not the same as either of the markers, co-purifies



A triad in human skeletal muscle consisting of a tubule of the T-system (T), formed by invagination of the plasma membrane, sandwiched between two terminal cisternae of the sarcoplasmic reticulum (SR). Depolarization, transmitted internally from the fibre surface by the T-system, triggers a transient release of calcium from the SR into the cytoplasm which in turn activates contraction of the myofilaments (M).

M.J. Cullen

with both of them.

While these experiments may mean that dystrophin is associated with the triads, it is not clear what form that association might take. For those with a morphologist's eye, a picture may be worth a thousand gels, and localization at the level of the electron microscope is eagerly awaited. Unfortunately, the very low abundance of dystrophin is likely to make this difficult unless it is highly concentrated at discrete sites.

The triads mediate a rapid increase in the intracellular level of free calcium. Because several lytic enzymes within cells are activated by calcium, it has been suggested^{8,9} that necrosis of muscle fibres results from an uncontrolled rise of calcium concentration within the cell. Although Hoffman *et al.* speculate² that the absence of dystrophin may somehow disrupt calcium homeostasis and cause the activation of proteases and phospho-

lipases, much work needs to be done before the details of this process are elucidated. If the absence of dystrophin is the primary cause of muscle-fibre degeneration in DMD, learning why this is so, however challenging, should tell us much about how the internal environment of muscle fibres is regulated. Understanding why the lack of dystrophin has such different effects on boys and mice may ultimately suggest ways of treating the many boys still dying from DMD. □

1. Hoffman, E.P. *et al.* *Cell* **51**, 919–928 (1987).
2. Hoffman, E.P. *et al.* *Nature* **330**, 754–759 (1987).
3. Hoffman, E.P. *et al.* *Science* **238**, 347–350 (1987).
4. Bulfield *et al.* *Proc. natn. Acad. Sci. U.S.A.* **81**, 1189 (1984).
5. Torres, L.F.B. & Duchon, L.W. *Brain* **110**, 269–299 (1987).
6. Wood, D.S. *et al.* *New Engl. J. Med.* **316**, 107–108 (1987).
7. Hammonds, R.G. *Cell* **51**, 1 (1987).
8. Cullen, M.J. & Fulthorpe, J.J. *J. neurol. Sci.* **24**, 179 (1975).
9. Mokri, B. & Engel, A.G. *Neurology (Minneapolis)* **25**, 1111–1120 (1975).

Clarke R. Slater is in the Muscular Dystrophy Group Research Laboratory, Newcastle General Hospital, Westgate Road, Newcastle upon Tyne NE4 6BE, UK.

History of science

H. Moseley and the Nobel prize

J.L. Heilbron

HENRY MOSELEY, born 100 years ago on 23 November 1887, established in 1913 the doctrine of atomic number by measurements on the characteristic X-ray spectra of the elements. Like many of his age and class (he came from a well-off academic family), he rushed to enlist at the outbreak of the First World War. Despite appeals from teachers and colleagues that he would be more useful at home, Moseley forced himself into the army engineers. His death in combat in Gallipoli, on 10 August 1915, was noticed on both sides; “*ein schwerer Verlust für die Naturwissenschaften*”, “a matter of great regret”, “*une mort glorieuse*” (see my book, *H.G.J. Moseley, The Life and Letters of an English Physicist*, Univ. California Press, 1974). This journal carried an obituary by his professor, Ernest Rutherford, who took the opportunity to berate the War Office for so “striking [an] example of the misuse of scientific talent”.

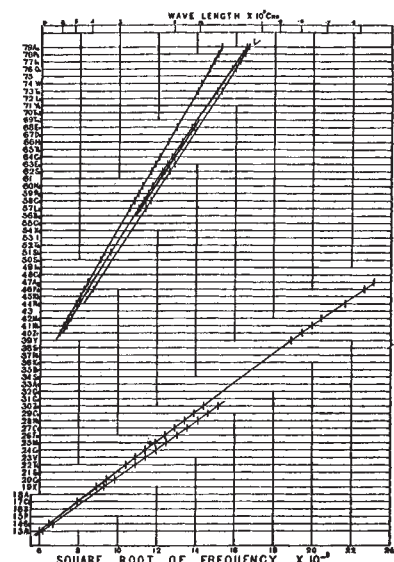
In 1912, M. von Laue and collaborators showed that crystals can diffract X-rays, and W.L. and W.H. Bragg demonstrated interference between X-rays reflected from a crystal. The following year, in Rutherford's laboratory in Manchester, Moseley perfected the Braggs' technique with C.G. Darwin and adapted it to measurements of the high-frequency line spectra emitted by atoms exposed to the broadband X-rays from a discharge tube. These spectra, like optical spectra, are characteristic for each atomic element, but in contrast can be expressed mathematically fairly simply. Thus, the frequency of the most intense lines (the K

lines) is roughly proportional to $(Z-1)^2$, where Z is the atomic number of the element. Moseley confirmed this for 10 metals from calcium to zinc. He later showed the similar relation for L X-rays for much of the periodic table while at Oxford (see graph). Where no known element fit the formula, he inferred that chemists had missed an opportunity. His method not only demonstrated the fundamental importance of the idea of atomic number, but also reliably revealed where elements remained to be discovered.

Had Moseley lived, he would probably soon have received the Nobel prize for physics or chemistry. He was nominated for both in 1915 by Svante Arrhenius, the most influential member of the committee on the Nobel physics prize of the Royal Swedish Academy of Sciences. The physics prize for 1914 had been reserved;



Henry Moseley — a wasted talent.



A Moseley diagram showing the relationship between X-ray frequency and atomic number (from Phil. Mag. 27, 703; 1914).

in 1915 the Academy decided to give no propaganda advantage to either warring camp by awarding that year's prize to the Braggs and the reserved one from 1914 to von Laue. No such solution was available in 1916, when nominees for the physics prize included several future winners from belligerent nations (Einstein, Perrin, Planck and Stark). The Academy again reserved the prizes in 1917 despite the renewed candidacy of Einstein, Planck and Stark in physics and Nernst in chemistry. The reservation became permanent in chemistry, but not so in physics. In 1918 the physics prize for 1917 was awarded to C.G. Barkla “for his discovery of the characteristic Röntgen radiation of the elements”.

Barkla was the *locum tenens* for Moseley. Barkla received only one nomination, from Rutherford, who argued that the discovery of characteristic X-rays was second in importance only to the discovery of their diffraction, which the Academy had already rewarded. Furthermore, as Rutherford observed, Barkla had inferred from his measurements important information about the number of electrons in an atom. The judges in Stockholm were thus invited to conclude that Barkla had made some progress toward the idea of atomic number using the same sort of X-rays Moseley was to use. In deciding to adopt Rutherford's suggestion, Arrhenius's committee considered that the value of Barkla's work, then a decade or more old, had been newly demonstrated by recent work — that is, by Moseley's work — in X-ray spectroscopy. It is difficult to resist the inference that, had it not been for the war, Moseley would have received, perhaps jointly with Barkla, the Nobel prize in physics for 1917. □

J.L. Heilbron is at the Office for the History of Science and Technology, University of California at Berkeley, California 94720, USA.

Dynamics of disease

Complex cardiac rhythms

Leon Glass

NOWADAYS, if one develops an abnormal heart rhythm, a visit to a cardiologist is in order. The cardiologist has received extensive training in identification of cardiac arrhythmias and has some idea of which are dangerous and how to treat them, but has little background in mathematics. However, if the report of Chialvo and Jalife on page 749 of this issue¹ is followed to its logical conclusion, the mathematical training of cardiologists will have to be upgraded. Or else a new specialist, a dynamicist, will have to assume a place amongst the other '-ists' of medicine.

Chialvo and Jalife applied a periodic stimulus to an *in vitro* preparation of sheep Purkinje fibre. This tissue, which forms part of the cardiac conduction system, is excitable but does not generate a spontaneous oscillation. At slow rates of stimulation, using large-amplitude stimuli, each stimulus is conducted, leading to a 1:1 conduction. By changing the stimulation strength and periodicity, it is possible to generate a large number of different conduction rhythms. These rhythms are characterized by two numbers N and M , where N is the number of stimuli and M the number of conducted responses in a repeating sequence ($N:M$ locking). In one experiment, Chialvo and Jalife measured the ratio of M/N as a function of increasing stimulation strength. As stimulation strength increases, keeping stimulation period constant, the conduction ratio M/N increases with plateaux at $1/4$, $1/3$, $1/2$ and 1. The resulting function resembles a mathematical function called a Cantor function or, more picturesquely, a "devil's staircase"². In a Cantor function between any two plateaux with ratios m/n and M/N , there is another one with ratio $(m+M)/(n+N)$. Therefore, there is an infinite number of steps leading to a staircase only the devil could climb. But the widths of the successive steps get very narrow, and are difficult to observe experimentally.

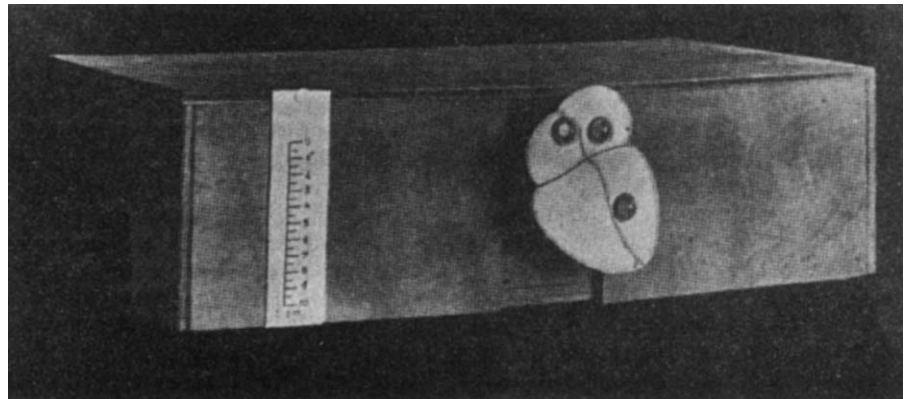
The rhythms observed in this experiment¹ seem to be analogous to the cardiac arrhythmias called AV heart block. In AV heart block cardiac excitation is generated normally by the sinus pacemaker but is blocked near the atrioventricular (AV) node which forms a conducting pathway between the atria and the ventricles. AV heart block is characterized by two integers which give the ratio of the number of atrial to ventricular excitations. The devil's staircase can also be observed in experimental data from intact human heart. The atria of patients are

periodically electrically stimulated using an intra-atrial catheter. The stimulation strength is kept constant, while the period of the stimulation is decreased³. At fast stimulation rates, heart block develops, and a devil's staircase provides a good description of the function representing the dependence of the conduction ratio on the stimulation frequency.

These recent results^{1,3} relate to pioneering studies by the early electrocardiologist Mobitz, who developed a graphical technique to determine conduction through the AV node⁴. The basic idea is that the conduction time through the node

Chialvo and Jalife also observe¹ other, more exotic, rhythms which are much harder to characterize and interpret. These include rhythms in which there is an alternation of latency from the stimulation to the conducted impulse, and complex aperiodic dynamics. Readers of News and Views will be familiar with the concept that the transition from periodic to aperiodic 'chaotic' dynamics often follows 'universal' transition sequences⁵. The examples in Chialvo and Jalife's paper provide further evidence that concepts from nonlinear mathematics are relevant in cardiac physiology.

The use of nonlinear mathematics to study complex dynamic phenomena in cardiology is in contrast to the currently popular biophysical methods used to characterize the various ionic currents and channels underlying cardiac activity. Similar dynamic behaviour can be found



In 1928, van der Pol and van der Mark developed an electronic model of the heart¹². The model was based on coupled nonlinear oscillators, which were constructed by charging capacitors that discharge through a neon tube once a threshold voltage is reached. There were three such oscillators corresponding to the sinus node, the atria and the ventricles. Spread of excitation in the heart model could be observed from the flashing neon tubes. By altering the frequency and coupling of the oscillators, it was possible to simulate many cardiac arrhythmias. In current jargon, the observed changes in the rhythms are called bifurcations. The study of bifurcations of periodically forced oscillators has had a deep impact on pure and applied mathematics and engineering, but cardiologists do not usually analyse arrhythmias using mathematical or electronic models.

depends on the recovery of the node following the last conducted impulse. Once this recovery function is known, it is possible to predict the rhythms that arise as a function of stimulation frequency. Mobitz recognized that many of the observed rhythms of heart block could be accounted for in this fashion. Although he did not discuss Cantor functions, his theoretical idea leads to such a functional relationship. Mobitz's argument seems to have been lost in modern cardiology teaching: I have never seen it reproduced in any cardiology textbook even though it seems to me to be a crucial insight. Mobitz's idea has been rediscovered periodically and has appeared in the research literature^{5,6}, but only recently in an appropriate mathematical context^{3,7,8}. Van der Pol, who was a contemporary of Mobitz, was also interested in the mechanisms of AV heart block, but approached the problem from a somewhat different perspective (see figure).

in all excitable tissue, be it the heart, brain, gut, ureter or even a non-living chemical soup that can propagate excitation¹⁰. It is not necessary to know the details of the ionic mechanisms underlying the activity to understand the complex rhythms that are supported by the excitable system. The ramifications of this perspective extend to many areas of biology and medicine¹¹.

What will make believers of even the most sceptical will be the use of nonlinear dynamics in medicine to help decrease morbidity and mortality. Although this is not yet possible, there is reason for optimism. One of the main causes of death is a sudden, unexpected, fatal arrhythmia in the heart. This is a problem because clinicians are not always able to distinguish dangerous from benign arrhythmias. Nonlinear mathematics may ultimately give a much deeper understanding of the dynamics of the complex arrhythmias that are frequently encountered clinically.

This in turn may help in the classification of the arrhythmias and identification of the most dangerous. Understanding the effects of drugs is not only a problem of pharmacology (what channel is affected?), but also of nonlinear dynamics (how does this change in ionic properties lead to novel dynamic organization of the heart?).

Both cardiology and nonlinear mathematics deal with the classification of complex dynamics. It is not remarkable that techniques and observations in one discipline are relevant in the other, but rather that both fields have been so independent over the years. □

Pharmacology

5-HT₃ receptors in the brain?

Philip Bradley

THE paper by Kilpatrick, Jones and Tyers on page 746 of this issue¹ presents the first experimental evidence for the existence of the 5-HT₃ subtype of receptors for 5-hydroxytryptamine (5-HT or serotonin) in the mammalian brain. This result could have important, exciting implications for the development of a new class of drugs acting on the brain.

It has long been known that 5-HT is a neurotransmitter in the central nervous system and that it is involved in various physiological functions such as sleep, control of blood pressure, temperature regulation, sexual activity and the perception of pain, as well as playing an important role in psychiatric conditions, for example, disturbances of affective behaviour such as depression and anxiety states. Much of the evidence for the involvement of 5-HT in these conditions comes from knowledge of the pharmacological properties of the drugs which are used in their treatment.

That there is more than one type of receptor for 5-HT was first suggested by Gaddum in 1954 and described in more detail² in 1957 when the two subtypes of receptor were designated 'M' and 'D'. Unfortunately, the classification of 5-HT receptors has since become confused, partly by the use of different terminology for the receptor subtypes and, perhaps more importantly, by the use of different criteria for classification. Recent proposals^{3,4} for the classification of 5-HT receptors are based on functional criteria, and it seems this classification is being used, although there are alternative proposals⁵.

It is generally accepted that there are three different subtypes of receptor for 5-HT and, according to the classification based on functional criteria³, they are called 5-HT₁-like, 5-HT₂ and 5-HT₃, although the 5-HT₁ receptor has been

1. Chialvo, D. & Jalife, J. *Nature* **330**, 749–752 (1987).
2. Mandelbrot, B. *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982).
3. Shrier, A. *et al.* *Circulation* **76**, 1196–1205 (1987).
4. Mobitz, W. *Zeit. f. d. Ges. Exp. Med.* **41**, 180–237 (1924).
5. Decherd, G.M. & Ruskin, A. *Br. Heart J.* **8**, 6–16 (1946).
6. Levy, M.N. *et al.* *Prog. Cardiovasc. Dis.* **16**, 601–613 (1974).
7. Keener, J.P. *J. Math. Biol.* **12**, 215–225 (1981).
8. Glass, L. *et al.* *Ann. N. Y. Acad. Sci.* **504**, 168–178 (1987).
9. Wolf, A. *Nature* **305**, 182–183 (1983).
10. Winfree, A.T. *When Time Breaks Down: The Three-Dimensional Dynamics of Electrochemical Waves and Cardiac Arrhythmias* (Princeton University Press, 1987).
11. Koslow, S.H. *et al.* (eds) *Ann. N. Y. Acad. Sci.* **504** (1987).
12. van der Pol, B. & van der Mark, J. *Phil. Mag.* **6**, 763–775 (1928).

Leon Glass is in the Department of Physiology, McGill University, Montreal, PQ, Canada H3G 1Y6.

further subdivided on the basis of data from binding studies. In the central nervous system the 5-HT₁-like receptor is a presynaptic, or autoreceptor, which controls the release of the transmitter from presynaptic nerve terminals. The 5-HT₂ receptor, which seems to correspond to Gaddum's D receptor, is present in the brain and is probably the principal postsynaptic receptor for 5-HT; it mediates the so-called 5-HT behavioural syndrome such as the 'head twitch' and 'wet dog shakes' in rats. The 5-HT₃ receptor, which corresponds to Gaddum's M receptor, has until now been found only on peripheral neurons where it mediates depolarization or release of transmitter. Kilpatrick *et al.*¹ now produce evidence suggesting that the 5-HT₃ receptor exists in the brain.

But there is need for caution. In the past there has been a tendency to assume too readily that a binding site is identical to a functional receptor; it is now recognized that this is not always the case. What Kilpatrick *et al.* have done is to produce evidence for the existence of a binding site in the brain, which shows the appropriate responses to drugs that are selective for peripheral 5-HT₃ sites, both agonists and antagonists, exactly as if it is a 5-HT₃ receptor. However, the only correlation the authors demonstrate between affinity for the binding site and pharmacological potency is for a peripheral 5-HT receptor on the rat vagus nerve (their Fig. 2b on page 748). Thus, the final proof that this 5-HT₃ binding site is a functional receptor is still awaited, and the authors go too far both with their title and in their statement "Here we report direct evidence for the existence of 5-HT₃ receptors in rat brain tissue and their distribution . . .". In my view, they should have confined themselves to the term 'binding site' in presenting their otherwise excellent data.

The final demonstration of the existence of functional 5-HT₃ receptors in the brain must come from studies of the central responses to drugs selective for this site. The drug used by Kilpatrick *et al.* in their binding studies, GR 65630, has not been shown to produce any behavioural effects. The closely related compound, GR 38032F, together with MDL 72222 and ICS 205-930, all of which are selective 5-HT₃ antagonists in the peripheral nervous system, and all of which compete for the binding of radioactively labelled GR 65630 in brain, have been reported to produce behavioural effects in rats and primates. However, the currently available data are all contained in abstracts of papers presented at meetings or in short reports which neither provide sufficient detail nor have been fully scrutinized.

What is needed is fully documented and detailed study of the central actions, if any, of GR 65630 and related compounds before the concept of 5-HT₃ receptors in the brain can be fully accepted. Such studies must be carried out with extreme care, first, to ensure that the effects seen are not caused by some peripheral action of the drug which then affects the central nervous system indirectly; and second, that they are not caused by an effect of the drug on some other neurotransmitter system.

The drug ketanserin, for example, is selective for 5-HT₂ receptors but some of its effects result from an action at α -adrenoceptors. It is possible, then, that GR 65630 interacts with neurotransmitter systems in the central nervous system other than 5-HT, despite the fact that ligands for other neurotransmitter receptors were shown to be inactive in the binding studies. If, as suggested by the brief reports⁷ of the behavioural effects of GR 38032F, this drug has anxiolytic properties, and if its pharmacological properties are similar to those of GR 65630, a new class of anti-anxiety drugs may appear. But caution is again needed, particularly in interpreting data from animal tests in terms of possible therapeutic actions in humans⁸. The question that remains is "what are the behavioural effects of the agonists and antagonists of 5-HT₃ receptors?" □

1. Kilpatrick, G.J., Jones, B.J. & Tyers, M.B. *Nature* **330**, 746–748 (1987).
2. Gaddum, J.H. & Piccarelli, Z.P. *Br. J. Pharmac.* **12**, 323–328 (1957).
3. Bradley, P.B. *Neuropharmacology* **25**, 563–576 (1986).
4. Saxena, P.R. *et al.* *Trends pharmac. Sci.* **7**, 424–428 (1987).
5. Green, J.P. *Trends pharmac. Sci.* **8**, 90–94 (1986).
6. Green, A.R. *et al.* *Neuropharmacology* **22**, 573–578 (1983).
7. Jones, B.J., Oakley, N.R. & Tyers, M.B. *Br. J. Pharmac.* **90**, 90P (1987).
8. File, S.E. *Trends Neurosci.* **10**, 461–463 (1987).

Philip Bradley is Emeritus Professor of Pharmacology at the University of Birmingham Medical School, Vincent Drive, Birmingham B15 2TJ, UK.

Evolution

Human bodies of evidence

Bob Martin and Paul H. Harvey

IN 1924, Raymond Dart discovered the first remains of *Australopithecus*. By 1953, his imagination, seeded by the association of australopithecine with other animal bones at various South African sites, led to his claim that our predecessors "were carnivorous creatures, that seized living quarries by violence, battered them to death, tore apart their broken bodies, dismembered them limb from limb, slaking their ravenous thirst with the hot blood of victims and greedily devouring livid writhing flesh" (*Int. Anthropol. Linguistic Rev.* 1, 209; 1953). Subsequent analysis of material from one of those sites revealed that australopithecines were prey rather than predators. This is just one particularly graphic over-enthusiastic interpretation of human fossil remains that was eventually rejected after sober scientific analysis. Representatives from different disciplines met recently in an attempt to provide a more secure basis for elucidating early human lifestyles*.

In the face of the relatively sparse fossil record of human evolution, scenes of murder, ritual slaughter and even cannibalism have repeatedly been built on flimsy evidence. Negative reactions to such sensationalism have tended to overshadow the steady progress that has been made in the systematic study of the biological life stories of our early relatives. Partly because of the introduction of new techniques, possibilities now exist for the reconstruction and interpretation of overall life-history patterns of early human populations.

Perhaps the most revolutionary contribution to the recognition of life-history patterns in fossil and archaeological specimens is emerging from microscope studies of teeth and bone. A. Boyde and T. Bromage (University College, London) and L. Lanyon (Royal Veterinary College, London) gave compelling accounts of how hard tissue can now be analysed to reveal details of the life-history of its original possessor (see figures). Incremental lines in teeth and bone can indicate age at death (see Bromage, T.G. & Dean, M.C. *Nature* 317, 525; 1985), whereas disruptions of the normal growth pattern can yield information on events such as birth, weaning, illness and possibly age at first breeding. Further, the resorption of bone, sometimes followed by repair, provides details of special stresses, for example those experienced by a lactating mother on an inadequate diet.

Other work discussed at the meeting demonstrates the value of combining broad comparative studies with detailed investigations of individual species and populations. We have explored the implications of scaling studies of life-history parameters across mammalian species (see Harvey, P.H. & Zammuto, R.M. *Nature* 315, 319; 1985 and Martin, R.D. & MacLarnon, A.M. *Nature* 313, 220; 1985). Such studies yield general principles that can be applied to the interpretation of individual species. For instance, there is a close correlation between age at first breeding and breeding lifespan in natural populations, irrespective of the effects of body size. Similarly, after size

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Tandem scanning reflected light microscope image deep to fractured surface of longitudinally split *Sivapithecus* molar tooth, 13 Myr before present, showing the daily incremental lines, or cross-striations, of the enamel prisms, and the circa-septan regular variety of the brown striae of Retzius. Tracor TSM Leitz 25; 0.75 oil immersion objective. (Courtesy of Alan Boyde.)

effects have been removed, the maximum reproductive potential of primates is typically higher for species living in savannah and secondary growth forests than in those from tropical rain forests where the environment is more predictable (Ross, C. J. *Zool., Lond.*; in the press). Among primates, as in mammals generally, age at first breeding has the greatest influence on overall reproductive potential and this particular life-history variable is thus worthy of special attention.

Detailed long-term observations of non-human primates in the wild, particularly macaques and chimpanzees, provide a fine-grained perspective both on the dynamics of life-history patterns and on variability among individuals. On the one hand, such studies underline the need for caution in the interpretation of small fossil samples; on the other, they provide a more reliable basis for making inferences. Rhesus macaques have figured prominently in providing baseline data of this

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Back-scattered electron image, diamond micro-milled, longitudinal section through a human molar tooth showing the cross-striations of the enamel prisms. Contrasts in this image are caused by density variations in the surface layer. Black and white bars are at intervals of 10 μ m. (Courtesy of Alan Boyde.)

kind, notably because of the long-term study of a free-ranging colony on the island of Cayo Santiago, Puerto Rico, first established in 1938 and monitored since 1956. D. Sade (Northwestern University) used his long experience of this colony to illuminate various aspects of life-history patterns. Computer-simulation studies based on long-term records from the colony confirm the theoretical expectation that age at first breeding is a key determinant of the rate of reproductive turnover. Interbirth intervals are of secondary importance, whereas longevity has relatively little influence.

The skeletons of some macaques that lived and died at Cayo Santiago, including at least one complete social group, have been preserved. As demonstrated by J. de Rousseau (Northwestern University), these skeletons are a valuable resource that will eventually permit full matching of actual life-history data (as the individuals were observed throughout their lives) with the messages to be read from teeth and bones. In a parallel development, M.E. Morbeck (University of Tucson) and A. Zihlman (University of Santa Cruz) are studying skeletons of the well-known chimpanzees from Gombe Stream Reserve in Tanzania that were the subject of Jane Goodall's project, which has now extended over more than 25 years (see Goodall, J. *The Chimpanzees of Gombe*; Belknap, Harvard, 1986).

Studies of human populations, both archaeological and living, add a further dimension to the picture. Excavations of mummified remains from cemeteries of a mediaeval Nubian community (D. Van Gerven, University of Colorado), supported by limited historical documentation, provide an astonishingly complete picture of individual life histories. Skilful detective work reveals not only a demographic profile of the population but also precise information on causes of mortality, particularly the effects of weaning stress among infants. Complementary

*Primate Life History and Evolution Wenner-Gren Foundation Conference 104, organized by M.E. Morbeck and J. de Rousseau. Cabo San Lucas, Baja, Mexico, 10-18 October 1987.

data from modern human populations can elucidate the costs and timing of reproduction in specific relation to interbirth intervals. N. Peacock (University of California, Los Angeles) reported on a field study of the Ituri pygmies from Zaire. This study, which incorporates the novel approach of measuring steroid hormone levels in samples of saliva, shows long interbirth intervals of 4 years or more. These seem to require some mechanism following on from the initial hormonal suppression of ovulation resulting from suckling. The energetic costs of locomotion involved in subsistence activities, combined with the burdens of carrying both a growing infant and the products of foraging, seem to be the determining factors. This has direct relevance for the interpretation of reproduction in relation to subsistence activities in early hominids.

Neanderthals were used as a test case for the practical application of these various approaches, and for exploration of their future potential. Fossils identified as Neanderthals constitute the largest sample (about 300 individuals) that is available for any group before the emergence of anatomically modern humans, and several immature individuals are known. Discussion animated by C. Stringer (Natural History Museum, London) and E. Trinkaus (University of New Mexico) led to some immediate benefits for the interpretation of Neanderthal life-history parameters, with the promise of more to come.

To date, only one apparently post-menopausal Neanderthal woman has been found and Neanderthals seem to have died relatively young as a rule. The prediction from comparative mammalian evidence, that under stable population conditions age at first breeding should also have been relatively earlier than in modern humans, remains to be tested. Various lines of comparative evidence now tend to exclude the suggestion that Neanderthals differed from modern humans in their gestation period (Trinkaus, E. *Curr. Anthropol.* **25**, 509; 1984). It seems likely, however, that they differed in other aspects of their developmental biology, notably during the early postnatal period. Unfortunately, precise ageing often still requires sectioning of teeth (Boyde), and museum curators are generally reluctant to release material. Non-invasive technology may eventually come to the rescue, but in the meantime a small loss of material is surely justifiable to allow absolute ageing of specimens. This will bring about the kind of revolution in anthropology that absolute dating brought about in palaeontology. □

Bob Martin is at the Anthropological Institute and Museum, University of Zurich-Irchel, CH-8057 Zurich, Switzerland; Paul H. Harvey is in the Department of Zoology, University of Oxford, Oxford OX1 3PS, UK.

Earth sciences

Similarities at fluid boundaries

Steve Thorpe

SIMILARITY, whether physical or biological, between different natural systems or components of a system, provides insight into their mechanics and, if expected, reassurance that understanding is soundly based. (It is perhaps more exciting if similarity is not found where it is expected, for then we have to refine the hypotheses.) The atmosphere-hydrosphere system of the Earth has boundary layers, where fluid meets solid earth or where one fluid (air) meets another (water). These layers are generally subject to the most extreme changes in physical properties, they contain by far the highest concentrations of biological organisms, and they are thus most prone to damage by pollution. How well can we estimate the consequences of such changes? Careful measurements by J.M. Brubaker, reported on page 742 of this issue¹, demonstrate the similarity between the night-time surface convective layer in lakes and the daytime surface convective regime in the lower atmosphere above land. These data improve confidence in predictive modelling studies.

The use of lakes for comparative studies with the ocean was enthusiastically advocated by Sir John Murray at the end of the last century and led eventually to the Bathymetric Survey of the Scottish Freshwater Lochs², completed in 1910. The two particular areas in which comparative measurements were thought to be useful were the phenomena of freezing and phosphorescence. (The latter still deserves consideration.) The most valuable scientific advance made by the Survey was in E.M. Wedderburn's work^{3,4} on the temperature and circulation of lakes, especially the discovery⁵, with E.R. Watson, of the 'temperature seiche', or standing internal wave, in Loch Ness. Temperature is the most easily measured physical quantity in liquids, using electrical measuring and recording devices such as platinum resistance thermometers. These were used to record³ rapid fluctuations in Loch Ness as early as 1903, and thermistor chains were used⁶ in 1952.

Brubaker's work can be seen as a further development in this tradition. His observations were made at a reservoir in Western Australia rather than a natural lake, using a freely rising microstructure probe to measure temperature. The project was part of a programme designed to develop a dynamical model of the reservoir, which could be used to 'manage' it. Turbulent fluxes and their effects in the upper layers are vital parts of the model, as in small-scale models of the ocean or in models of the lower layers of the

atmosphere. The similarity in temperature fluctuations between the 5-m-thick surface layer in the reservoir and the typically 1,000-m-thick atmospheric boundary layer adds to previous observations⁷⁻⁹ of dynamical similarity between oceanic and atmospheric boundary layers. It also boosts our confidence in developing similar models to predict variability or, for example, the spread of pollutants.

So where do we stand now? It seems that the effects of turbulent convection can be described in terms of conditions that prevail at the surface where convection occurs. The resulting distributions of passive, non-motile organisms, particles or chemical tracers in all the analogous regimes can be inferred from those in one. The internal waves generated at the edges of the convective regions by the buffeting of convective plumes, and the plumes themselves, will have similar structures. There is valuable information to be gleaned from atmospheric observations about how intensively the ocean must be sampled to obtain reliable averages.

But not everything is similar: humidity must be accounted for in the atmosphere, salinity in the ocean. Fortunately Brubaker made his measurements on an exceptionally cold night of strong convection (resulting in a dense, chilling radiation fog), but which was also calm: no waves on the water surface affected the convection or contributed to turbulence. Their effect is yet to be properly assessed¹⁰.

Although nocturnal convection in the ocean may often occur in calm weather, rarely will deep-winter mixing not be accompanied by wind and waves, which can also be responsible for vertical motions which resemble convection (for example, Langmuir circulation¹¹) about which too little is known. Waves must lead to turbulence close below the surface which is not dynamically similar to that in the atmospheric boundary layer over rigid terrain, and further comparisons in the upper-ocean boundary layer, and near the surface of lakes, may provide yet more insight into the basic phenomena. □

1. Brubaker, J.M. *Nature* **330**, 742-745 (1987).
2. Murray, J. & Puller, L. *Bathymetric Survey of the Scottish Freshwater Lochs* (Challenger Office, Edinburgh, 1910).
3. Wedderburn, E.M. *Trans. R. Soc. Edinb.* **45**, 407 (1907).
4. Wedderburn, E.M. & Watson, W. *Proc. R. Soc. Edinb.* **19**, 619-647 (1909).
5. Watson, E.R. *Geol. J.* **24**, 430-437 (1904).
6. Mortimer, C.H. *Proc. Int. Ass. Limnol.* **12**, 66 (1955).
7. Dillon, T.M. *et al. Nature* **290**, 390-396 (1981).
8. Shay, T.M. & Gregg, M.C. *Nature* **310**, 282-285 (1984).
9. Shay, T.M. & Gregg, M.C. *Nature* **311**, 84 (1984).
10. Thorpe, S.A. *Nature* **318**, 519-522 (1985).
11. Leibovich, S.A. *Rev. Fluid Mech.* **15**, 391-427 (1983).

Steve Thorpe is Professor of Oceanography at Southampton University, Southampton SO9 5NH, UK.

Translational control

Developments in development

Mike Clemens

STUDIES on the mechanisms by which higher cells regulate their rates of protein synthesis are currently enjoying something of a revival, as became apparent at a recent meeting*. This interest can be attributed partly to the cloning and sequencing of some of the eukaryotic polypeptide chain initiation factors, and partly to the development of antibodies against these factors. As a result, it is becoming clear that there are multiple ways in which overall rates of translation can be regulated in a response to external stimuli. In no system is this more obvious than the favourite cells of developmental biologists, the eggs of sea urchins and frogs.

For several years dogma has held that the state of profound translational inactivity characteristic of unfertilized sea urchin eggs or immature oocytes of the South African frog *Xenopus laevis* is caused by 'masked' maternal messenger (m) RNA which is somehow prevented from interacting with the abundant ribosomes present in these cells¹. Following fertilization or progesterone-induced maturation, respectively, this mRNA becomes more available to the translational machinery and the rate of protein synthesis increases. It is now apparent that this is not the whole story, as several recent reports make clear. It now seems that there is activation of at least three initiation factors in addition to the use of previously masked mRNA in these systems. Apart from the intrinsic interest of these findings, the identification of these regulatory mechanisms is significant for our understanding of translational control in somatic mammalian cells subjected to environmental stresses such as heat shock or serum deprivation.

Many of the new results stem from the establishment of active cell-free protein-synthesizing systems which are capable of initiation of translation. Protein synthesis in lysates from unfertilized eggs of the sea urchin *Lytechinus pictus* is not increased when exogenous mRNAs alone are added, but can be stimulated by addition of initiation factor eIF-2 (responsible for binding the initiator methionyl-transfer RNA (Met-tRNA) to the ribosome) or the guanine-nucleotide exchange factor GEF (which catalyses the recycling of eIF-2 between successive rounds of initiation) (M. Hille, University of Washington, Seattle)². However, this stimulation occurs only in the presence of added mRNA, indicating that neither eIF-2 nor

GEF can activate endogenous masked maternal mRNAs.

Another initiation factor, eIF-4F, also stimulates translation in this system and this is additive with the effect of GEF. Factor eIF-4F (see table) is a three-subunit protein involved in the recognition of the 5' cap structures of mRNAs and in the binding of mRNAs to ribosomal subunits that already carry Met-tRNA. The conclusion from these results in the sea urchin cell-free system is that not only is much of the maternal mRNA masked (presumably by specific proteins) but that the steps of Met-tRNA binding and mRNA binding are also defective (even for 'active' mRNAs) and therefore contribute to the state of translational dormancy characteristic of unfertilized eggs.

To what extent are these restrictions of initiation-factor activity reversed when

on eIF-2 α could be similar to those characterized in rabbit reticulocytes or interferon-treated mammalian cells⁴. In the case of eIF-4E, there is some evidence for the involvement of protein kinase C because, at least in HeLa cells, phorbol ester treatment stimulates phosphorylation of this protein and purified kinase C will modify the factor *in vitro* (R. Duncan, University of California, Davis). Changes in the activity of a kinase or phosphatase could provide a basis for the effects of an inhibitor of eIF-4F activity which has been identified in R. Jagus's laboratory (University of Pittsburgh) in extracts of unfertilized sea urchin eggs and which disappears on fertilization⁵. This possibility remains to be tested.

A similar limitation in binding of mRNA to ribosomes might occur in immature *Xenopus* oocytes, where initiation complexes containing Met-tRNA bound to small ribosomal subunits can be shown to accumulate *in vitro* (T. Patrick and J. Pain, University of Sussex). This result is consistent with a report that injection of the mRNA binding initiation factor eIF-4A (also one of the subunits of eIF-4F)

Properties of polypeptide chain initiation factor eIF-4F

Subunit	Mass	Synonym	Function
p26	24–28K	CBP I, eIF-4E	Binding of 5' caps of mRNAs ATP-dependent mRNA unwinding and binding of mRNA to ribosomes
p45	45–46K	eIF-4A	
p220	210–220K	—	Required for function of eIF-4F in translation of capped mRNAs

protein synthesis is stimulated after fertilization of sea urchin eggs, and by what mechanisms? Results from A. Lopo's laboratory (University of California, Riverside) indicate that phosphorylation of the M_r 28,000 (28K) subunit of eIF-4F (also called eIF-4E) is stimulated by fertilization. Conversely, the 36K α -subunit of eIF-2 becomes partially dephosphorylated under these conditions (C. Gallo and J. Hershey, University of California, Davis). It has been known for many years, from studies on rabbit reticulocytes, that the activity of eIF-2 can be inversely correlated with the extent of phosphorylation of its α -subunit³, and the fertilization-induced change is therefore consistent with reversal of a deficiency in eIF-2 activity. In the case of eIF-4E, it remains to be determined whether phosphorylation increases the affinity for mRNA cap structures. If this turns out to be the case, such an effect might increase the rate of efficiency of mRNA binding to ribosomes after fertilization by improving the process of cap recognition.

Clearly, because the phosphorylation states of eIF-2 α and eIF-4E change in opposite senses, different protein kinases and phosphatases probably modify these two substrates. The enzymes that act

stimulates translation of endogenous mRNA in immature oocytes but not in progesterone-matured oocytes⁶. But in this system no translational inhibitor can be detected.

Taken together, these results point to a multifactorial mechanism for the activation of protein synthesis during development in which more than one initiation factor is modified, in parallel with the unmasking of mRNA by a process that remains elusive. In addition, fertilization of sea urchin eggs and progesterone stimulation of *Xenopus* oocytes induce changes to parameters of the intracellular environment, such as pH, that may also contribute to the optimization of translation.

All these findings are significant for our understanding of translational regulation in other systems. In mammalian cells, initiation factors equivalent to those in the sea urchin system undergo covalent modifications under conditions that modulate translation. The α -subunit of eIF-2 and the 25K (eIF-4E) subunit of eIF-4F, for example, become phosphorylated and dephosphorylated, respectively, in response to heat shock or nutrient/serum deprivation of cultured mammalian cells⁷. These changes correlate with losses of activity which can be corrected by

*Translational Control Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 16–20 September 1987.

adding back the purified factors to cell-free systems prepared from cells subjected to such physiological stresses^{8,9}.

Furthermore, modulation of intracellular pH and ionic conditions (especially Ca^{2+}), such as occurs when quiescent cells are stimulated by growth factors or serum, is capable of regulating translation in mammalian cells (E.C. Henshaw and R. Panniers, University of Rochester Cancer Center). There are therefore many aspects in common between translational control in eggs and the responses of somatic cells to growth-regulatory conditions. □

1. Raff, R.A. in *Cell Biology, A Comprehensive Treatise* (eds Prescott, C.D.M. & Goldstein, L.) 107–136 (Academic, Orlando, 1980).
2. Colin, A.M. *et al. Devl Biol.* **123**, 354–363 (1987).
3. Pain, V.M. *Biochem. J.* **235**, 625–637 (1986).
4. Pestka, S., Langer, J.A., Zoon, K.C. & Samuel, C.E. *A. Rev. Biochem.* **56**, 727–777 (1987).
5. Huang, W.-I., Hansen, L.J., Merrick, W.C. & Jagus, R. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6359–6363 (1987).
6. Audet, R.G., Goodchild, J. & Richter, J.D. *Devl Biol.* **121**, 58–68 (1987).
7. Duncan, R., Milburn, S.C. & Hershey, J.W.B. *J. biol. Chem.* **262**, 380–388 (1987).
8. Panniers, R., Stewart, E.B., Merrick, W.C. & Henshaw, E.C. *J. biol. Chem.* **260**, 9648–9653 (1985).
9. Clemens, M.J. *et al. J. biol. Chem.* **262**, 767–771 (1987).

Mike Clemens is in the Department of Biochemistry, St George's Hospital Medical School, London SW170RE, UK.

Cosmology

Can nucleons close the Universe?

K. A. Olive

THE standard Big Bang model of the early Universe is tested observationally at two very different epochs. First, the observation of the 3-K microwave background discovered by Penzias and Wilson¹ probes back to the time when the age of the Universe was about 10^5 yr (today the age of the Universe is about 10^{10} yr). Second, Big Bang nucleosynthesis (see, for example, ref. 2), the era when the light nuclei D, ^3He , ^4He and ^7Li were produced, occurred about 1 min after the Big Bang. The microwave background radiation confirms the existence of our hot past and provides invaluable information regarding the overall isotropy of the Universe. Because nucleosynthesis probes back to much hotter temperatures (about 10^{10} K) and hence much earlier times, however, the abundances of the light nuclei are sensitive to many microphysical processes predicted in theories of high-energy physics. The effects of the transition from quarks and gluons (the constituents of nuclear matter) to hadrons (normal nuclear matter) before nucleosynthesis have now been investigated^{3–6} and contrasted with standard-model predictions.

The fact that the standard model of nucleosynthesis provides excellent agreement with the observational determination of the light-nuclear abundances allows the theories of particle physics to be constrained. A prime example of a nucleosynthesis constraint is the restriction on the number of light neutrino types^{7,8}. One additional neutrino type beyond the known three begins to produce an overabundance of ^4He . Any more beyond that would cause a conflict between theory and observations.

Agreement for all of the numerical abundances, which span ten orders of magnitude ($\text{D}/\text{H} \approx 10^{-5}$, $^3\text{He}/\text{H} \approx 10^{-5}$, $^4\text{He}/\text{H} \approx 0.08$ and $^7\text{Li}/\text{H} \approx 10^{-10}$), occurs only for a limited range of the ratio η of the number of baryons (protons and

neutrons) to the number of photons, $3 \times 10^{-10} < \eta < 6 \times 10^{-10}$. In addition, because the density of photons, n_γ , is known (by measuring the temperature of the microwave background), we can relate η to the ratio Ω_b of the total mass density in baryons, ρ_b , to the critical density, ρ_c , needed to close the Universe: $\Omega_b = \rho_b/\rho_c = 3.6 \times 10^7 \eta h_0^{-2} (T_0/2.7 \text{ K})^3$, where $H_0 = 100 h_0 \text{ km s}^{-1} \text{ Mpc}^{-1}$ is the present value of the Hubble parameter and T_0 is the present temperature of the microwave background. The maximum value of Ω_b , obtained with $\eta = 6 \times 10^{-10}$, $h_0 = 0.4$ and $T_0 = 2.8 \text{ K}$, is 0.15, thus excluding the possibility that the Universe is closed (so that its expansion asymptotically stops) by ordinary matter.

Recently there have been some interesting proposals for altering the standard scheme for nucleosynthesis^{3–6}. These use the effects of a phase transition between normal, hadronic matter and quark–gluon matter which is thought to have existed about 10^{-5} s after the Big Bang. In particular, depending on the largely unknown details of this transition, baryon density fluctuations can arise as a result of the difficulty in producing a baryon with mass $m_b \approx 1 \text{ GeV}$ at a temperature $kT = 200 \text{ MeV}$ (k is Boltzmann's constant). If indeed baryon production is suppressed in the hadron phase, an excess will appear in the quark phase which, as the transition proceeds, occupies less and less space, thus resulting in a baryon density fluctuation.

It has been known for some time that baryon density fluctuations could affect Big Bang nucleosynthesis, but the effects were rather minimal. Applegate and Hogan³ suggested, and now with Scherrer⁴ show, that in this primordial medium, the mean free paths of neutrons and protons are different, leading to a segregation of neutrons and protons. The Universe after the transition is then divided into neutron-rich and neutron-

poor regions. The effect of this inhomogeneity on nucleosynthesis can be great.

Because the abundances of D, ^3He and ^4He as a function of η are largely monotonic, a suitable choice of transition parameters can hold the average abundance of each of these nuclei constant, as η is increased to the point where $\Omega_b = 1$. But ^7Li abundance does not change monotonically as a function of η . In the standard model of nucleosynthesis, there is a minimum for the abundance of ^7Li at $\eta = 3 \times 10^{-10}$ for which the fraction of the mass of the Universe in ^7Li , $X(^7\text{Li}) = 4 \times 10^{-10}$. Either increasing or decreasing η results in a higher ^7Li abundance⁵. Thus, although Alcock *et al.*⁶ can replicate standard Big Bang nucleosynthesis results for D, ^3He and ^4He and $\Omega_b = 1$, they find that the abundance of ^7Li ranges from 3×10^{-8} to 10^{-7} , which should be compared with observational determinations of ^7Li abundance.

Because ^7Li can be both produced and destroyed in other astrophysical sites such as stars and cosmic-ray spallation, it is not a trivial task to extract the primordial ^7Li abundance from the observations. Indeed, observations of ^7Li in population II stars imply $X(^7\text{Li}) \approx 6 \times 10^{-10}$ and in population I stars $X(^7\text{Li}) \approx 7 \times 10^{-9}$, although there are arguments⁹ that suggest that population II abundances more accurately represent the primordial abundances. In any case, both measures of ^7Li abundance fall well short of the values produced in nucleosynthesis calculations based on inhomogeneities induced by the quark–hadron transition.

I would like to stress that this should not be construed as a negative result. Instead, these results can be used to place constraints on the quark–hadron transition as has been done by Reeves¹⁰. Using both measures of $X(^7\text{Li})$, he places limits on the unknown parameters of the quark–hadron transition, such as the baryon-number contrast and the fractional mass in the two phases. These constraints imply that η lies in the restricted range $1 \times 10^{-10} - 6 \times 10^{-10}$, or $\Omega_b \leq 0.15$ as before, and thus exclude the possibility of closing the Universe with nucleons. □

1. Penzias, A.A. & Wilson, R.W. *Astrophys. J.* **142**, 419 (1965).
2. Borgeard, A. & Steigman, G. *A. Rev. astr. Astrophys.* **23**, 319–378 (1985).
3. Applegate, J.H. & Hogan, C.J. *Phys. Rev. D* **31**, 3037–3045 (1985).
4. Applegate, J.H., Hogan, C.J. & Scherrer, R.J. *Phys. Rev. D* **35**, 1151–1160 (1987).
5. Sale, K.E. & Mathews, C.J. *Astrophys. J.* **302**, L1–L4 (1986).
6. Alcock, C., Fuller, G.M. & Mathews, G.J. *Astrophys. J.* **320**, 439–447 (1987).
7. Steigman, G., Schramm, D.N. & Gunn, J.E. *Phys. Lett.* **66B**, 202–204 (1977).
8. Yang, J., Turner, M.S., Steigman, G., Schramm, D.N. & Olive, K.A. *Astrophys. J.* **281**, 493–511 (1984).
9. Kawano, L., Schramm, D.N. & Steigman, G. *Astrophys. J.* (in the press).
10. Reeves, H. *Proc. Int. Sch. Phys. Enrico Fermi* (in the press).

K.A. Olive is in the School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, USA.

Adaptive and non-adaptive suicide in aphids

SIR — McAllister and Roitberg¹ recently reported evidence for 'adaptive suicide'² in pea aphids (*Acyrtosiphon pisum*). The aphids were said to have three strategies of escape from predators: backing-up, running, and falling from the food plant. Parasitized aphids from dry regions used the falling strategy more often than expected and this was put forward as suicidal behaviour.

The authors attempted to exclude the possibility that the apparently suicidal behaviour was a direct maladaptive effect of the parasite: they compared the behaviour of aphids from coastal and from dry habitats, where the fitness consequences of falling were different. However, the experimental rationale relies upon the unfounded assumption that the two aphid populations were genetically homogeneous at loci not involved in anti-predator behaviour. Indeed, in a previous paper³, it is implicitly stated that aphids living in two areas, coastal and dry, "are considered different biotypes". Therefore, aphids from the dry habitat may simply be more susceptible to the direct effects of the parasite.

The aphids used were only parasitized by wasps from their own habitat. Therefore, the observed differences between aphid populations may have been a function of the parasites rather than the hosts. A standardized experiment using all four host/parasite combinations could be performed to examine this possibility.

The increased falling behaviour of parasitized aphids from hot regions may simply be a superior means of escape from predators. The authors give an imprecise indication of the fitness consequences of different escape strategies in each of the two aphid habitats. Moreover, they do not indicate whether the fitness consequences differ between parasitized and unparasitized individuals.

The conclusion that suicide is adaptive will apply only if it is under genetic control. The authors have not shown this. Crosses between the two aphid populations are needed.

Following evolutionary stable strategy theory⁴, it is likely that all three escape strategies have equal payoffs at equilibrium, despite having apparently different fitness consequences. Parasitized individuals could commit suicide by performing any escape strategy in excess. Perhaps the dropping strategy is most efficient for suicide in hot areas, but this does not explain the absence of suicidal behaviour in the coastal population. The fitness of parasitized individuals from the temperate coastal region would still be lowered if an excess chose to jump rather than perform an optimal mixture of strategies.

Why is suicide so inefficient? It is curious that the parasitized aphids committed suicide only when approached by a ladybird, or when escape pheromones had been detected. More obvious means of committing suicide would simply be not to eat or to fall from the food plant spontaneously. It was proposed that the suggested suicide mechanism evolved because it was an easy modification of an existing behavioural strategy. Even so, the proposed modification of behaviour is surprisingly slight.

Finally, it is not in the interest of parasites that their hosts should commit suicide. Natural selection on parasites may favour the subversion of the host's suicidal behaviour⁵. In any ensuing 'arms race', asymmetries of selection should favour the parasite.

In my opinion, the evidence of McAllister and Roitberg does not give convincing support to the 'host suicide hypothesis'. Their results are compatible with a number of other hypotheses, both adaptive and non-adaptive.

IAN TOMLINSON

Department of Genetics,
University of Cambridge,
219D Huntingdon Road,
Cambridge CB3 0DL, UK

1. McAllister, M.K. & Roitberg, B.D. *Nature* **328**, 797-799 (1987).
2. Smith Trail, D.R. *Am. Nat.* **116**, 77-91 (1981).
3. Roitberg, B.D. & Myers, J.H. *Can. Ent.* **111**, 515-519 (1979).
4. Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge University Press, 1982).
5. Dawkins, R. *The Extended Phenotype* (Freeman, Oxford, 1982).

SIR—I wish to question the reasoning behind a recent paper by McAllister and Roitberg¹, intended to demonstrate apparently suicidal behaviour in parasitized pea aphids from British Columbia. Pea aphids were chosen because of their parthenogenetic mode of reproduction and low dispersal rate — both factors conducive to verification of the 'host suicide hypothesis'², which holds that a host whose reproductive potential has effectively been reduced to zero by a parasitoid may yet enhance its inclusive fitness³ by actively increasing the probability of death, so preventing maturation of the endoparasite and preventing its spread among closely aggregated clonal kin.

McAllister and Roitberg argue that pea aphids in a moist climate stand a reasonable chance of survival on the ground, whereas pea aphids from a hot, dry climate quickly die of heat stress and desiccation. In other words, under normal circumstances, an aphid from a dry climate would do better in terms of its individual fitness to remain on the leaf and back up when approached by a predator than to drop from the leaf — both of which are possible escape behaviours used by aphids⁴. If the host suicide hypothesis holds, then parasitized aphids from a dry

climate could enhance their inclusive fitness by increased dropping under the stimulus of a predator, while parasitized aphids from a moist climate would not increase the probability of death by dropping and so their escape behaviour would not alter when parasitized. The reported experiment seemed to confirm these predictions.

The difficulty I find with this argument is that, to increase its inclusive fitness, the host aphid need not commit suicide, but merely need isolate itself (along with its endoparasite) from its stationary aggregate of kin. That is, once the host has dropped, it makes no difference to the fitness of its kin (and hence its inclusive fitness) whether death on the ground is the result of this behaviour or not. In fact, as its aim is to increase the fitness of its kin relative to the fitness of non-kin⁵, it would do better to remain alive and so allow the parasitoid to mature, because mathematically the parasitoid is more likely to find its way to aggregates of non-kin than back to the original aggregate of kin.

On the basis of this 'host isolation hypothesis', I suggest that, as described in the report¹, there should have been increased dropping of parasitized dry-climate aphids, but, contrary to the reported findings, there should have been no difference between the behaviour of dry-climate aphids and moist-climate aphids. Both should have increased dropping behaviour when parasitized.

I am at a loss to explain the apparent success of the reported experiment. But, where a premise is in error, agreement between prediction and observation cannot be taken as confirmation of the original hypothesis.

BLAIR LATTA

97 Division Street,
Kingston, Ontario,
K7L 3M2 Canada

1. McAllister, M.K. & Roitberg, B.D. *Nature* **328**, 797-799 (1987).
2. Smith Trail, D.R. *Am. Nat.* **116**, 77-91 (1981).
3. Hamilton, W.D. *J. theor. Biol.* **7**, 1-52 (1964).
4. Brodsky, L.M. & Barlow, C.A. *Can. J. Zool.* **64**, 937-939 (1986).
5. Lewontin, R.C. *Sci. Am.* **239**(3), 213-230 (1978).

Translational control in homoeobox mRNAs?

SIR—The deduced sequences of several complete vertebrate homoeobox complementary (c)DNAs have recently been published. We have searched for regions conserved in genes from frogs¹⁻⁴, mice⁵⁻⁷ and humans^{8,9} and have found extensive amino-acid conservation of X1Hbox 2, HHO.cl, XHox-36 and Hox 1.1 throughout the whole coding region, such that they form a distinct family. The degree of similarity of X1Hbox 2 and HHO.cl suggests that these are the frog and human homologues of the same gene; in the mouse their homologue is called Hox 2.3¹⁰.



Conserved regions in the 5'-non-coding regions in homoeobox-containing genes. The initiator AUG is at the right-hand end of the figure. Vertical lines between pairs of sequences indicate identical nucleotides. Regions of high conservation are boxed. A short region conserved in the *Drosophila* gene *Deformed* is indicated. *Fushi tarazu* has a small stretch of nucleotides (TCTGATTTTGCTATATAT, approximately -100 upstream of the AUG) similar to the second and third boxed regions of X1Hbox 2 and HHO.c1 (not shown).

Most interestingly, we unexpectedly found extensive nucleotide conservations in the 5'-non-coding regions immediately preceding the initiator AUG, as shown in the figure. The conservation is maximal among the four genes mentioned above. For X1Hbox 2 (ref. 3) and HHO.c1 (ref. 9) it is 91 % (over 99 nucleotides), for Xhox-36 (ref. 2) and Hox 1.1 (ref. 7) it is 77 % (over 102 nucleotides), and for Hox 1.1 and X1Hbox 2 it is 77 % (over 100 nucleotides). Sequence conservations in this region occur in other genes: mouse Hox 2.1 (ref. 5) and Hox 1.3 (ref. 6) are 63 % conserved over 114 nucleotides.

Why are non-coding nucleotides conserved in many genes from such diverse organisms? Perhaps this merely reflects the common evolutionary origin of homoeobox genes. Alternatively, the sequences may have remained invariant because of strong evolutionary pressure to preserve a common function. The best argument in favour of this is provided by the homologues X1Hbox 2 and HHO.c1 (ref. 9). These two genes have been evolving for at least 350 million years, since the separation of amphibians and the mammalian ancestors. Their 5'-non-coding regions next to the AUG are 91 % conserved, but the 3'-non-coding regions show no sequence conservation.

In the coding region between the amino terminus and the homoeobox there is very little nucleotide similarity and even in the homoeobox regions, where the protein reading frame is under strong pressure to remain invariant, the conservation is only 81 %. We think the 5'-non-coding region is the most conserved not only because of a common evolutionary origin but because it carries an important biological function.

The conservations shown in the figure are not due to a translational reading frame, because in all sequences there are stop codons in all reading frames and frequent insertions that would produce frameshifts. As the 5'-leader conservations are closely associated with the initiation codon, an attractive hypothesis is that these sequences are involved in translational control. Translational control has been implicated in the expression of homoeobox genes in *Drosophila*: the best

example being the maternal gene *caudal* whose mRNA, although uniformly distributed in the egg, is translated only in the posterior of the embryo^{11,12}. The possibility of translational regulation has been shown for the frog gene X1Hbox 2 in an experiment in which deletion of most of the 5'-leader of a cDNA clone, leaving only 26 nucleotides in front of the AUG, stimulated *in vitro* translation of SP6 messenger RNAs over 20-fold by both reticulocyte and wheat-germ systems⁴. The proteins translated from both constructs were of the same size⁴, providing direct evidence that the 5'-conserved region is not translated.

THOMAS R. BÜRGLIN
CHRISTOPHER V.E. WRIGHT
EDDY M. DE ROBERTIS

Department of Biological Chemistry,
University of California,
Los Angeles, California 90024-1737, USA

1. Harvey, R.P., Tabin, C.J. & Melton, D.A. *EMBO J.* **5**, 1237-1244 (1986).
2. Condie, B.G. & Harland, R.M. *Development* **101**, 93-106 (1987).
3. Wright, C.V.E., Cho, K.W.Y., Fritz, A., Bürglin, T. & De Robertis, E.M. *EMBO J.* (in the press).
4. Fritz, A. & De Robertis, E.M. *Nucleic Acids Res.* (in the press).
5. Krumlauf, R., Holland, P.W., McVey, J.H. & Hogan, B.L.M. *Development* **99**, 603-617 (1987).
6. Odenwald, W.F. et al. *Genes Dev.* **1**, 482-496 (1987).
7. Kessel, M., Schulze, F., Fibi, M. & Gruss, P. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5306-5310 (1987).
8. Mavilio, F. et al. *Nature* **324**, 664-668 (1986).
9. Simeone, A. et al. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4914-4918 (1987).
10. Meijlink, F. et al. *Nucleic Acids Res.* **15**, 6773-6786 (1987).
11. Mlodzik, M. & Gehring, W.J. *Cell* **48**, 465-478 (1987).
12. MacDonald, P.M. & Struhl, G. *Nature* **324**, 537-545 (1986).

AIDS incubation period in male haemophiliacs

SIR—Ekert¹ concludes that there is no adequate explanation for the tenfold lower rate of conversion of infection with the human immunodeficiency virus (HIV) to clinical AIDS (acquired immune deficiency syndrome) in male haemophiliacs compared with other groups of HIV carriers. A possible explanation is that, unlike heterosexual male haemophiliacs, other groups of HIV carriers are exposed during sexual intercourse to the immunosuppressant effects of seminal plasma.

Human seminal plasma contains various immunosuppressive agents². In par-

ticular, seminal plasma may inhibit the normal immune response to a viral infection³. Prostaglandin E₂, a constituent of seminal plasma, has been shown *in vitro* to facilitate HIV replication⁴. Clinically, both the transmission⁵ and the neoplastic complications⁶ of HIV are associated with receptive anal intercourse. We have recently suggested that following sexual intercourse the suppression of the immune response to HIV by seminal plasma may be important in patients who are already HIV carriers³. Such an effect could hasten the development of clinical AIDS in HIV carriers and is consistent with the higher incidence of Kaposi's sarcoma⁶ and lymphomas⁷ in male homosexual HIV carriers compared with other HIV carriers, including haemophiliacs.

Thus, a longer incubation period and a lower incidence of clinical AIDS in haemophiliacs following HIV infection may be more apparent than real because of the shorter incubation period in other HIV carriers.

M.J. TURNER
J.O. WHITE
W.P. SOUTTER

Institute of Obstetrics and Gynaecology,
Royal Postgraduate Medical School,
Du Cane Road, London W12 0HS, UK

1. Ekert, H. *Nature* **329**, 494 (1987).
2. Alexander, N.J. & Anderson, D.J. *Fertil. Steril.* **47**, 192-205 (1987).
3. Turner, M.J., White, J.O. & Soutter, W.P. *Immun. Today* **8**, 258 (1987).
4. Kuno, S. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3487-3490 (1986).
5. Kingsley, L.A. et al. *Lancet* **i**, 345-349 (1987).
6. Marmor, M. et al. *Ann. intern. Med.* **100**, 809-815 (1984).
7. *Lancet* **i**, 193-194 (1986).

HIV vaccination and blood transfusion

SIR—With the increasing attention being given to the possibility of vaccination against human immunodeficiency virus (HIV)¹⁻³, we wish to raise the question of the impact of vaccination on blood donor screening.

Vaccination would produce seropositive individuals whose blood is reactive with HIV-antibody screening tests currently in use for blood donor screening. As even confirmatory tests could not establish whether a positive test was the result of

vaccination rather than infection, donations will be discarded before a definite answer is at hand. Certificates of vaccination, safeguarded against falsification, would help, but only if donors make the effort to carry them.

What happens if vaccination is not completely protective against infection with HIV, as in the case of chimpanzees⁴? In this case even the introduction of certificates would not guarantee that a vaccinated donor was uninfected.

In our opinion the maintenance of an adequate blood supply will be jeopardized by large-scale vaccination against HIV. How are we to avoid this problem?

DIETMAR FUCHS

ARNO HAUSEN

GILBERT REIBNEGGER

DIETHER SCHOENITZER

ERNST R. WERNER

MANFRED P. DIERICH

HELMUT WACHTER

*Institute of Medical Chemistry and Biochemistry,
Blood Transfusion Centre, and Ludwig Boltzmann Institute of AIDS Research,
University of Innsbruck,
A-6020 Innsbruck, Austria*

1. Zagury, D. *et al.* *Nature* **326**, 249–250 (1987).

2. Coates, A.R.M. *et al.* *Nature* **326**, 549–550 (1987).

3. Salk, J. *et al.* *Nature* **326**, 473–476 (1987).

4. Hu, S.L. *et al.* *Nature* **328**, 721–723 (1987).

Lead and children

SIR—In their article¹ “Lead and child development”, Davis and Svendsgaard refer to evidence that postnatal mental development of children may be adversely affected at (umbilical cord) blood-lead levels as low as 6–7 $\mu\text{g dl}^{-1}$. The work to which they allude has now been published in full². In the course of our research on placental element levels in relation to human fetal development³ we have measured umbilical cord blood-lead levels for 100 obstetrically normal births in Barnsley in the United Kingdom. The range was 6–39 $\mu\text{g dl}^{-1}$, and the arithmetic mean 14 $\mu\text{g dl}^{-1}$. It therefore appears reasonable to infer that most if not all of these ‘normal’ neonates had been exposed *in utero* to levels of lead sufficient to be associated with some postnatal mental maldevelopment.

We scarcely need to labour the social and educational implications of this.

In general, we find that placental lead levels relate highly significantly to birth-weight and other indices of prenatal development in a negative and dose-dependent manner, whereas lead levels in maternal and cord blood and maternal and neonate hair do not. We have, however, found that the levels of cadmium and zinc in placenta also correlate significantly with prenatal development, negatively and positively respectively³. So it may be that fetal exposure to these further two elements also needs to be considered in

the aetiology of human postnatal maldevelopment, and that placenta is a better indicator tissue than blood.

D. BRYCE-SMITH

*Department of Chemistry,
University of Reading,
Whiteknights, PO Box 224,
Reading, Berkshire RG6 2AD, UK*

N.I. WARD

*Department of Chemistry,
University of Surrey,
Guildford GU2 5XH, UK*

1. Davis, J.M. & Svendsgaard, D.J. *Nature* **329**, 297 (1987).
2. Bellinger, D., Leviton, A., Waternaux, C., Needleman, H., & Rabinowitz, M. *New Engl. J. Med.* **316**, 1037–1043 (1987).
3. Ward, N.I., Watson, R. & Bryce-Smith, D. *Int. J. Biosoc. Res.* **9** (1), 63–81 (1987).

Galactic chronology of thorium/neodymium

SIR—Butcher¹ has used his observations of thorium and neodymium lines in stars of different relative ages to determine the duration of nucleosynthesis over the lifetime of the Galaxy. He concludes that the age of the Galaxy is ≤ 9.6 Gyr for synthesis at a constant rate throughout the lifetime of the Galaxy. Here I compare Butcher's data with the galactic chronology I discussed in ref. 2. I conclude that my cosmochronological model based on ^{232}Th , ^{235}U and ^{238}U as chronometers is in agreement with Butcher's observations when the ages of the stars in his observations are scaled to my galactic age equal to $10.0 \pm 1.5(1\sigma)$ Gyr. My calculation is made using equation (1) (equation 38 in ref. 2).

This equation gives the ratio of abundance N divided by relative production rate P for any isotope A to any standard S . The mean lifetimes are designated by τ_A and τ_S respectively. I have generalized equation (38) by replacing Δ , the duration of nucleosynthesis before the Solar System formed, where necessary, with t , the duration for a star formed at time t after the start of nucleosynthesis in the Galaxy taken as $t = 0$. The right hand side of equation (1) is based on a chronological model for nucleosynthesis in the Galaxy with an early spike, S_E , followed by uniform nucleosynthesis, $1 - S_E$, as discussed by Meish and me³ in connection with the rapid neutron capture process (r-process).

In ref. 2 the calculation of r-process production for the isotopes of U and Th were updated and used in combination with abundances at $t = \Delta$ inferred from meteorite data corrected to the origin of

the Solar System $4.6 \pm 0.1(1\sigma)$ Gyr ago. Two ratios were available, $^{232}\text{Th}/^{238}\text{U}$ and $^{235}\text{U}/^{238}\text{U}$, so two free parameters were determined, namely S_E and Δ . The values found were $S_E = 0.17 \pm 0.02(1\sigma)$ and $\Delta = 5.4 \pm 1.5(1\sigma)$ Gyr. Consequently the age of the Galaxy was given as $t_0 = 4.6 \pm 0.1 + 5.4 \pm 1.5 = 10.0 \pm 1.5(1\sigma)$ Gyr. This value agrees within the quoted errors with the value I obtained², $t_0 = 12.4 \pm 2(1\sigma)$ Gyr, using the nuclear cosmochronology I originally developed with Burbidge, Burbidge and Hoyle⁵ and later perfected with Hoyle⁶.

Butcher's observations involve radioactive ^{232}Th ($\tau = 20.27$ Gyr) and the seven stable isotopes of Nd ($A = 142$ –146, 148, 150). Butcher determined the relative abundances of Th and Nd in stars at the present time t_0 , as a function of the time, t , of formation of the star. He designated this ratio as Th/Nd, in my notation equation (2).

If all of the isotopes of Nd had been produced in the r-process it would only be necessary to set $\tau_S = \infty$ in equations (1) and (2) taking Nd as the ‘standard’. However Howard *et al.*⁷ have analysed the production of the isotopes of Nd and have shown that the total element was produced 50 per cent in the r-process and 50 per cent in the s-process, the slow neutron capture process of ref. 5. In addition Sneden and Pilachowski⁸ have reported observations on old metal-poor stars which can be interpreted by omitting an early spike in s-process nucleosynthesis. Only uniform synthesis over the lifetime of the Galaxy is necessary. Thus in applying equation (1) to Th and Nd the value for S_E in the denominator is $0.17/2 = 0.085$ and $(1 - S_E)$ becomes 0.915. Equations (1) and (2) then yield equation (3).

In addition an overall normalization factor, f , is required in comparing Butcher's Th/Nd data with calculations using equations (1), (2) and (3) as $P(\text{Th})/(\text{Nd})$ cannot be accurately calculated. This comparison is shown in Fig. 1.

In Fig. 1 the ordinate shows the abundance ratios, Th/Nd, observed by Butcher¹ in 19 stars including the Sun with the uncertainties in the ratios quoted by him. The abscissa is the age, $t_0 - t$ in my notation, of the 19 stars assigned by Butcher on the basis of values he was able to ascertain from recent literature. I have divided Butcher's age scale by a factor of 2 to agree with my value, $t_0 = 10$ Gyr, with one exception. The point for the Sun is plotted

$$\frac{N_A(t)/N_S(t)}{P_A/P_S} = \frac{S_E \exp(-t/\tau_A) + (1 - S_E)(\tau_A/\Delta)[1 - \exp(-t/\tau_A)]}{S_E \exp(-t/\tau_S) + (1 - S_E)(\tau_S/\Delta)[1 - \exp(-t/\tau_S)]} \quad (1)$$

$$N_A(t_0)/N_S(t_0) = \exp(t/\tau_A - t_0/\tau_A - t/\tau_S + t_0/\tau_S) N_A(t)/N_S(t) \quad (2)$$

$$\frac{\text{Th}}{\text{Nd}} = \frac{0.170 + 3.1156 \times [1 - \exp(-t/20.27)]}{0.085 + 0.1694 \times t} \times f \exp[(t - 10)/20.27] \quad (3)$$

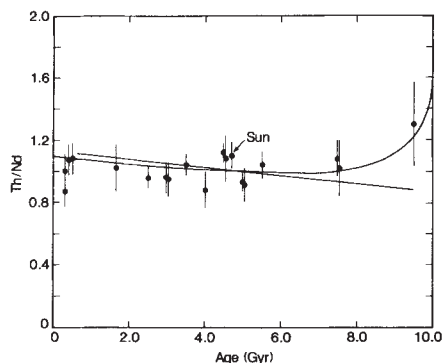


Fig. 1 The observations by Butcher¹ on the abundance ratio, Th/Nd, for 19 stars including the Sun plotted vs the age of the stars adjusted to the age of the Galaxy given in ref. 2. The curve is a theoretical calculation based on the nuclear cosmochronology described in ref. 2.

at the accepted age of 4.6 Gyr. The theoretical curve calculated as described above has been normalized to Th/Nd = 1 at the age of the Sun with $f = 1.40$. It will be noted that the curve fits the observations reasonably well. It must be conceded that Butcher's data can be fitted with other values of f , S_E and t_0 . This can be understood simply because the mean lifetime of ²³²Th is so long, about 20 Gyr. At the same time I think it fair to emphasize that the agreement shown in Fig. 1 is based on a chronological model² presented in the literature before Butcher's observations were published. That model used a range of mean lifetimes, 1.0 Gyr for ²³⁵U, 6.4 Gyr for ²³⁸U, as well as 20 Gyr for ²³²Th.

Faced with my timescale, the astute reader will question how the published ages for Butcher's stars can be high by a factor of 2 both by his calculations and by mine. The published ages come in the main from the time required for a star in a globular cluster to reach the main-sequence red giant turnoff point in the Hertzsprung–Russell diagram. I think that a possible solution of the problem has been given by Clayton⁹ and Willson, Bowen and Struck-Marcel¹⁰. These authors introduce early main-sequence mass loss for stars and the latter authors state, "We conclude that it is possible that no stars in our Galaxy are older than $7-10 \times 10^9$ yr old." If stars are more massive when they first form they convert their primordial hydrogen into helium very quickly and thus reach the main-sequence red giant turn-off point when their central hydrogen

is consumed much sooner than given by canonical calculations. I have summarized² many other observations that agreed with a Galaxy age of 10 Gyr.

I acknowledge correspondence on Butcher's observations with H.R. Butcher, D.D. Clayton, G.J. Mathews and D.N. Schramm and the receipt of preprints from Clayton, Mathews and Schramm. This work was partly supported by the NSF.

WILLIAM A. FOWLER

W.K. Kellogg Radiation Laboratory,
California Institute of Technology,
Pasadena, California 91125, USA

BUTCHER REPLIES — Professor Fowler's enthusiasm for my result is gratifying, especially so because I was once a student of his. Let me emphasize, however, that the correct treatment of s- and r-process abundances in models for the chemical evolution of the Galaxy is very uncertain. What seems to have been established¹ is that the ratio of these abundances does not vary more than the measurement error — the best to date being about 25 per cent — among stars of all ages, as long as the absolute heavy element abundance level is greater than about 3 per cent of its solar value. How this result may be reconciled with major ongoing stellar synthesis in an evolving galaxy is unclear. There do exist, of course, stars with even smaller metal abundances, and in these stars one does observe variations in the relative amounts of r- and s-process elements. One imagines that such stars are evidence for a period early in the Galaxy before much nucleosynthesis has occurred, but how they relate in time to other, indistinguishably old stars with higher abundances may only be guessed at. It is fortunate that for the timescale discussion, which uses stars with abundances high enough to show the thorium line, the problem of different synthesis possibilities for neodymium does not seem to matter.

HARVEY BUTCHER

Kapteyn Astronomical Institute,
Postbox 800,
9700 Av. Groningen,
The Netherlands

1. Lambert, D.L. *J. Astrophys. Astr.* **8**, 103–122 (1987).

SIR—There is no contradiction between Butcher's¹ Th/Nd ages for the oldest stars of about 11–12 Gyr and the results of stellar evolution, when the diffusion of helium is taken into account. The evolution of stars of low mass is accelerated by the diffusion of helium towards the centre^{2,3}. When this process is omitted, the ages are overestimated. At a nominal age of 16 Gyr as calculated by classical methods⁴ (omitting diffusion), the actual age^{2,3} is about 25 per cent less, or 12 Gyr. This reduces the evolutionary ages in Butcher's sample (minus at most one standard deviation) to lie within the range of his nucleochrono-

logical results. It is in fact quite encouraging that Butcher's ages are as young as they are; otherwise one would have to invoke some mechanism to slow the evolution of these stars. Meridional circulation could do so by mixing them if they were rapid rotators, but Butcher's own data belie this unless the stars rotate much more rapidly inside than on their surfaces or have spun down quite recently.

PETER D. NOERDLINGER

4702 Hayman Avenue,
La Canada, California 91011, USA

1. Butcher, H.R. *Nature* **328**, 127–131 (1986).
2. Noerdlinger, P.D. & Arigo, R.J. *Astrophys. J.* **237**, L15–L16 (1980).
3. Stringfellow, G.S., Bodenheimer, P., Noerdlinger, P.D. & Arigo, R.J. *Astrophys. J.* **264**, 228–236 (1983).
4. VandenBerg, D.A. *Astrophys. J. Suppl. Ser.* **58**, 711–769 (1985).

BUTCHER REPLIES — Noerdlinger brings to our attention one of several plausible mechanisms for reducing the maximum ages of stars as deduced from stellar evolution calculations. This idea has not yet been tested against the helioseismological data which are not available, so it is hard to see that it is clearly preferable to other age-reducing possibilities. Another suggestion that has received considerable discussion recently, proposes that solar-mass stars in reality were substantially more massive as they arrived on the main sequence, but then suffered a period of extensive mass loss. Such stars would then have burnt significant amounts of their core hydrogen very quickly, and would appear rather older than they really are¹. There might be observable signatures of this process, in the distribution of main-sequence stars versus mass for example, but at present no convincing evidence one way or the other has been presented to decide the matter. Finally, it has been recognized for some time that even the standard models used for computing stellar evolution can produce substantially younger ages than those reported as best estimates, by adjusting the detailed chemical composition, the description of convection and the distance scale, maximum ages as young as 10 Gyr may be obtained².

Determining which of these ideas may apply to the real world, and which are red herrings, may have to await the new generation of very large optical telescopes, when high-quality, high-resolution spectra and detailed asteroseismological data can be obtained for relatively faint stars.

HARVEY BUTCHER

Kapteyn Astronomical Institute,
Postbox 800,
9700 Av. Groningen,
The Netherlands

1. Willson, L.A., Bowen, G.H. & Struck-Marcell, C. *Commun. Astrophys.* **12**, 17–34 (1987).
2. Janes, K. & Demarque, P. *Astrophys. J.* **264**, 206–214 (1983).

1. Butcher, H.R. *Nature* **328**, 127–131 (1987).
2. Fowler, W.A. *Q. J. R. astr. Soc.* **28**, 87–108 (1987).
3. Fowler, W.A. & Meist, C.C. *Cosmological Process* (eds Arnett, W.D., Hansen, C.J., Truran, J.W. & Tsuruta, S.) 83–100 (VNU, Utrecht, 1986).
4. Fowler, W.A. *Proc. of the Rutherford Jubilee Int. Conf.* 640–676 (Heywood & Co Ltd, London, 1962).
5. Burbidge, E.M., Burbidge, G.R., Fowler, W.A. & Hoyle, F. *Rev. mod. Phys.* **29**, 547–650 (1957).
6. Fowler, W.A. & Hoyle, F. *Ann. Phys.* **10**, 280–302 (1960).
7. Howard, W.M., Mathews, G.J., Takahashi, K. & Ward, R.A. *Astrophys. J.* **309**, 633–652 (1986).
8. Sneden, C. & Pilachowski, C.A. *Astrophys. J.* **288**, L55–L58 (1985).
9. Clayton, D.D. *Astrophys. J.* **140**, 1604 (1964).
10. Willson, L.A., Bowen, G.H. & Struck-Marcell, C. *Commun. Astrophys.* **12**, 17–34 (1987).

Temporal and spiritual

C.J.S. Clarke

Time, The Familiar Stranger. By J.T. Fraser. *University of Massachusetts Press: 1987.* Pp. 389. \$24.95.*

It is not a new idea that time is multiple, that there is more than one structure to which the word refers; this explains why the discussion of time is so often fraught with confusion and paradox. But Fraser is the first to have developed the idea into a universal vision, condensed into the modest space of this book.

His temporal categories, identifying the different sorts of time appropriate to different levels of organization, are: atemporal, for the most fundamental constituents of the universe; proto-temporal, for massive particles; eotemporal, for macroscopic non-living objects; biotemporal for life; nootemporal for thinking organisms; and perhaps socio-temporal for humanity. The time of our immediate experience, with which Fraser starts his analysis, is the nootemporal one, and each lower form has less structure. Pedagogically this is a wise approach. It is easy for the physicist to forget that, though the time of special relativity is structurally much simpler than the "devouring time" of Shakespeare's sonnets (of which Fraser is agreeably fond), it is for that very reason more remote and harder to understand.

A deeper lesson which the physicist can learn from this approach lies in the possibility that the time at one level may not necessarily be derivable from that at the previous level simply by adding structure, such as adding directionality to a continuum. There may in fact be no simple relation between them. It is becoming increasingly clear, for example, that, if a successful unified theory of fundamental interactions emerges from the current investigation of string-theories, it will involve a 'space-time' whose relation to the space and time of observation is indirect and has to be worked out through a suitable theory of measurement.

Much of the appeal of the book lies in its wealth of charming detail: in the information that the *Farmer's Almanac* gives the year-number of the current Roman Indiction ("useful for farmers who pay taxes to the emperor"); in Proustian vignettes of encounters under beech-trees; in the assessment of political ambition by the

remark that "English people stand for office, Americans run for it".

Only once does this civilized style become heated, when the author contemplates thermodynamic treatments of 'the arrow of time'. Although he has made it clear that the central meaning of the arrow of time must be sought in nootemporality (specifically, in memory and anticipation), it remains the case that there are important links between this and circumstances that are not time-symmetric

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

obtaining at lower levels. Nootemporality could not evolve in a universe where the temporal direction of electromagnetic radiation is not related to the expansion of the universe in the way that is the case; nor would it make sense if charges were to radiate to the nootemporal past. Physicists have much to say about these links, between time at one level and physical processes in time at another level, and by using phrases such as "knee-jerk metaphysics" Fraser does less than justice to the achievements of Reichenbach, Gold, Sciama and others.

Almost nothing of interest is excluded from the scope of this book, and contention is therefore to be expected. There is much religious material, mostly well handled. But it was surprising, for instance, to see such a deeply Platonic thinker as Rumi enlisted at the end of Chapter 3 in support of an evolutionary view, and my surprise increased when the experience of the Mevlevi dervishes, stemming from him, was compared to

"the mass hysteria of screaming, weeping, and fainting generated by the rock 'n' roll in the 1960s". This forms part of a section arguing that experiences of timelessness are regressions to lower levels of temporality rather than glimpses of a higher one, but no reference is made to any of the main Western mystical writers.

The section with the biggest impact is the final dark anticipation of the emergence of a new level of temporality that has become possible with the establishment of a single present for the whole of humanity, linked together through global communications. Acknowledging that he is working more on intuition than cast-iron logic, Fraser points to many signs of "preadaptation" for this next evolutionary step. Our established attitudes to time weaken. The sense of history is lost as people and governments concentrate on the present and on short-term goals. The

rhythm of the calendar is lost. Skills of analysis and interpretation become the preserve of a small élite, because "In the coevolution of people and their computers, the most significant step has not been that computers have become user-friendly, but that people have become computer-friendly".

As the machineries of multinational capitalism and state-capitalism become increasingly alike, the Utopia offered to humanity is of "an orderly world, one without problems, a society where the needs of the citizen for bodily comfort are taken care of by magic invisible powers, where his demands for mental stimulation are satisfied by simple questions and undisturbing answers, and where disturbing individuals are joyfully disposed of". The new

global society, if it comes into being, unchecked by any external constraint but troubled by internal tensions and shortages of resources, will be prone to wild instability which it may or may not survive. We are now at the transitional phase, called by Fraser "the Anthill Threshold".

We might reply that social evolution is a matter for conscious choice, and no path is inevitable. But in practice humanity has muddled along with little reflection, until we now find ourselves at the point where the resources needed to avert global starvation are being devoted to the means of global incineration. Fraser reveals a chilling third alternative, the Anthill which will be ours if we opt for short-term technological solutions with no long-term goal. As well as setting the context for future work on time, this book will add to our understanding of the choice before us for the future direction of humanity. □

C.J.S. Clarke is a Professor in the Faculty of Mathematical Studies, University of Southampton, Southampton SO9 5NH, UK.

*In Britain the book is available from Eurospan, 3 Henrietta Street, London WC2E 8LU. Price is £20.

Sampling the deep Earth

D. K. Bailey

Mantle Xenoliths. Edited by Peter H. Nixon. Wiley:1987. Pp.844. £100, \$215.

TRADITIONALLY, the main concern of geology has been with the Earth's crust, constituting less than 1 per cent of the planet's volume. Most of the Earth (84 per cent) is made up of the underlying mantle, and in recent years there have been international efforts to explore its composition (as distinct from its geophysical characteristics). Samples of the solid mantle are available in two forms: large slices (measured in kilometres) driven through the crust in episodes of mountain building, and small fragments brought to the surface by high speed volcanic eruption (xenoliths). Although the latter are haphazard samples torn from the sides of the volcanic conduits, they are especially valuable because there has been little chance of alteration during their transport to the surface. An assemblage of xenoliths from one eruptive centre can provide a profile through the upper mantle to depths as great as 250 kilometres. The evidence is fragmentary, in every sense of the word, but these are the only direct samples available from the major portion of our planet, and any attempt to produce a survey of existing knowledge must be welcome.

The book is divided into two parts. The first is devoted to regional studies and is aimed at providing a comprehensive review of the worldwide distribution of mantle xenolith locations in five geographic regions (lithosphere plates). After an introductory overview by the editor, several chapters are devoted to each plate, each chapter dealing with a different part of the plate. Coverage is variable, largely because of the paucity of information from particular regions, for example South America and India, but there are some puzzling omissions, such as West Germany in a fulsome coverage of the rest of Europe. A separate section devoted to mantle xenoliths within the ocean basins would have been useful, even though these may appear (at the moment) to be similar to some continental examples. Any deficiencies, however, are more than balanced by the gathering together of dispersed information, and invaluable surveys of hitherto poorly known or unevenly described regions. China has many geological marvels and the chapters here presage more wondrous things to come.

All in all, Part 1 is a splendid array of information, given coherence by the excellent overview and tables provided by the editor. One almost wishes he had

presented his own comprehensive atlas, but even so this format constitutes a major contribution in this complex field.

The first part of the book is an entity and could stand alone. Part 2 is devoted to "Principles, Processes and Special Studies" and, as its title suggests, is less coherent than the first part, suffering from a lack of linkages and general overview by the editor. The structural rationale is hard to discern, and the brief editorial introduction does not provide much guidance in spite of an amusing cartoon showing the chapter numbers on a schematic section through the Earth. Many of the individual contributions are excellent and some clearly relevant to principles and processes. Why some topics were singled out for special treatment, however, is less clear, especially when others, such as eruptive mechanisms, or melting processes, lack separate coverage.

In a book of this size it would certainly have helped the general reader to have had a chapter (or appendix) in which the classification of mantle fragments, in terms of modal mineralogy and fabrics, was clearly set out. The glossary is not sufficient. Naturally, in a compendium of

50 chapters (78 contributors) there is a wide variation not just in style but even in attack. Uniformity of terminology is thus all the more necessary: terms like 'asthenosphere', for instance, are difficult enough without the spectre of undefined and variable usage.

Many data have been brought together in well-constructed tables, which together with the maps provide a mine of information. The abundant colour plates not only inform but delight the senses, and the unfamiliar reader will readily appreciate one of the fascinations of these rocks.

There are a good many books dealing with the mantle, and mantle processes; some are restricted in theme and many are conference volumes, and as a consequence coverage of the subject is dispersed and uneven. This position has now been substantially rectified by the immense efforts realized in this book, which is a testimony both to the editor's fortitude, and to his insight, born of long experience in the field of mantle xenoliths. □

D. K. Bailey is a Professor in the Department of Geology, University of Reading, Reading RG6 2AB, UK.

Social climbing

James W. Valentine

Evolution and Escalation: An Ecological History of Life. By Geerat J. Vermeij. Princeton University Press: 1987. Pp.527. \$47.50, £31.70.

THIS book is about adaptation and adaptations, especially those that arise as a result of interactions among species. The uses to which adaptations are devoted are certainly most easily studied in living organisms, but the success of particular sorts of adaptations in contributing to the dominance or longevity of lineages is best assessed in the fossil record. With fossil evidence one can commonly follow the origin, distribution and persistence of adaptations over long spans of geological time. Vermeij is a marine biologist, experienced in functional interpretations of morphology, and he knows the fossil record well. Here he traces some of the consequences of morphological adaptations and generalizes upon the patterns they display.

The book is divided into four parts. The first sets out a theoretical framework for the study of morphological aspects of adaptation. Although it is accepted that these adaptations evolve through fitness differentials, there is little or no understanding of the genes or gene systems underlying the morphologies involved, and therefore Vermeij employs a vocabulary suited to phenotypic rather than to

genotypic aspects of adaptation.

For example, adaptive traits (termed "aptations" regardless of whether they originally evolved to serve their present functions) arise through the action of selective agencies termed "hazards". The probability that an individual will survive an encounter with a hazard is a measure of the "effectiveness" of the appropriate aptation, while the "adaptive gap" is a measure of unsuccessful encounters; thus these terms substitute for "fitness" and "selective intensity" as used in the conventional evolutionary vocabulary. Within an evolving lineage, adaptive improvement occurs as effectiveness increases and the adaptive gap narrows. If a hazard is intensified, the response may be a further enhancement of the associated aptations, and "escalation" is said to have occurred.

The second part is entitled "The Acquisition of Resources", and the third part (the longest, and the heart of the book) "The Evolution of Armor and Locomotion". In both of them Vermeij goes into predatory and anti-predatory aptations in detail; competition is deemed to be important also, but is given less attention, perhaps because the morphology of competition is not so clear. In Part Three, aptations developed in organisms with distinctive hard-part ground-plans — bivalved organisms, articulated organisms and so forth — are explored group by group. Most attention is focused on molluscs, the dominant armoured phylum in the seas today. The fourth part summarizes the evidence marshalled earlier in the book, and discusses the conditions favourable

for escalation, the times during the Phanerozoic when escalation was particularly prevalent, and the possible role of escalation in extinction and vice versa.

Vermeij has written a delightful book, brimming with data and replete with intriguing correlations and models of adaptation, all deftly presented. His vocabulary lends a distinctive flavour to the exposition, one that enables the reader to conjure up images of selective processes that involve differential mortalities rather than differential natalities, though it can serve both.

As the examples pile up, they come to form an impressive body of evidence to support Vermeij's main contention — that adaptations, especially those devoted to interspecies interactions, have increased during the course of time. Among the more important trends of escalation identified are increases in the capacity of marine predators to drill and break shells, with concomitant increases in the incidence and expression of armour in a wide variety of invertebrate groups; increases in the depth of burrowing or boring of prey and predators; and increased mobility in many groups. In Vermeij's words, "living species . . . are not necessarily better able to cope with their enemies than ancient ones were with theirs, but the biological surroundings have in an absolute sense become more challenging." I began as a sceptic on this matter but now can accept that the evidence does indeed favour escalation as being responsible for many of the differences between early and late morphologies within the principal invertebrate clades.

When it comes to the relevance of escalations to the major features in the history of life, Vermeij enters into an arena where there are numerous competing hypotheses, and his evidence for widespread escalation does not appear to bear heavily upon the outcome. It is striking that escalation has produced only relatively marginal changes. Considering how rapidly evolution can act at times, the differences between the armour of marine invertebrates of the Ordovician and those of today do not seem worth half-a-billion years of evolution. Vermeij recognizes this, and is at pains to point out that extinctions do not ordinarily spare the more escalated forms, which may indeed be more at risk. Can there be a sort of feedback at work here, with the short-term advantages of adaptation to biological hazards producing long-term disadvantages (perhaps owing to specialization rather than escalation *per se*) that soon undo them, geologically speaking?

Vermeij repeatedly states that, if it should turn out that higher-level processes override the sort of adaptation that he discusses, then his conclusions may not be valid. But don't believe that for a moment. As he well understands, the situation is far more complicated and his evidence and the contentions stemming therefrom are clearly important, regardless of the significance of other events. Anyone wishing to ponder what is important and what is not must read this book □

James W. Valentine is a Professor in the Department of Geological Sciences, University of California, Santa Barbara, California 93106, USA.

Bacteria in close-up

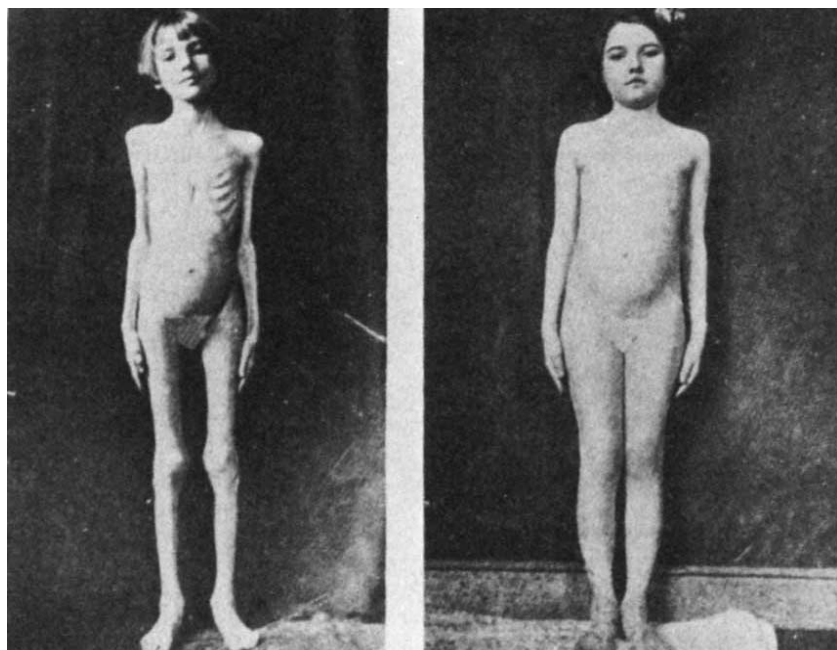
B. M. Wilkins & R. H. Pritchard

Escherichia coli and *Salmonella typhimurium*: Cellular and Molecular Biology, Volumes 1 & 2. Editor-in-chief Frederick C. Neidhardt. American Society for Microbiology, 1913 I St, NW, Washington, DC 20006:1987. Pp.1,654. Hbk \$100; (\$76 to ASM members); flexicover \$90 (\$66 to ASM members).

THE objective of these two volumes, in the words of the editors, is to provide for the first time a comprehensive treatment of the molecular and cellular biology of *Escherichia coli* and *Salmonella typhimurium* that would be a logical first resource for anyone seeking information about these organisms. With few reservations we think that this objective has been achieved, and achieved at a price for the flexicover edition that makes the books an exceptional bargain.

The two volumes are divided into six parts, with a total of 104 chapters written by about 150 authors who are recognized authorities. The total number of literature citations must be nearly 20,000. Despite these encyclopaedic dimensions and its multi-author make-up, the treatise is remarkably free from the overlap and the imbalance that frequently mars this kind of enterprise. It is equally pleasing to be able to report that the majority of authors have succeeded in writing what their editors required of them: "thoughtful and narrative reviews" as opposed to mere compilations of data and references.

Some chapters, certainly, are mainly and necessarily for reference. The most recently published linkage maps of both species are reproduced with minor modifications but with the bibliographies extended from earlier editions to obviate the need to refer back to them. Also provided is an updated version of pedigrees and genotypes of most of the widely used strains of *E. coli* that were previously published in *Bacteriological Reviews* in 1972, and a new edition of the gene-protein index previously published in *Microbiological Reviews* in 1983. New and helpful reference chapters are provided on the location of natural insertion sequences and on selectable markers. There is also a useful map of known transposon insertions. The remaining chapters deal with specific or general topics arranged in the six parts. Volume 1 contains "Molecular Architecture and Assembly of Cell Parts", and "Metabolism and General Physiology". Volume 2 contains "Genome and Genetics", "Regulation and Gene Expression", "Growth of Cells and Cultures", and "Ecology, Evolution and Population Structure". These titles give a



Miracle made manifest — before and after pictures, from 1922, of one of the first patients to be treated with insulin. The pictures, thought to be too indelicate for lay viewing at the time, are reproduced from *The Discovery of Insulin* by Michael Bliss, which has only now been published in Britain (by Macmillan, price £25). The book appeared in North America in 1982 and was reviewed in *Nature* 303, 261 (1983).

ADVERTISEMENTS

This is a great book,
and an exciting book. . .
readable, worth
reading and
enlightening.

Sir Karl Popper

Cosmology, Physics & Philosophy

by
Prof Ben Gal-Or

A Master Piece!
*Indian Journal of
Physics*

. . . shines a welcome
light in some dark
corners of science.
New Scientist

2nd Edition
Springer-Verlag, 1987
175 5th Ave.,
N.Y., 10010.

Reader Service No.7

Announcing J. T. Fraser's
**TIME, THE FAMILIAR
STRANGER**

No part of existence is more elusive
and mysterious than time. Written
in an engaging narrative style,
Fraser considers time in the every-
day world, in science and the
humanities - a fascinating history of
the idea and experience of time.

December 1987, 328pp, illus.,
Cloth £16.95



THE UNIVERSITY OF
MASSACHUSETTS PRESS
3 Henrietta Street,
London WC2E 8LU

Reader Service No.6

Copies of articles from
this publication are now
available from the UMI
Article Clearinghouse.

**UMI
Article
Clearinghouse**

Mail to: University Microfilms International
300 North Zeeb Road, Box 91 Ann Arbor, MI 48106

good indication of the general areas that
are covered.

The very specialized reader may not
find that his or her particular area of in-
terest is covered — there is no information
on catabolism of aromatic compounds, for
example. Perhaps more surprising is the
absence of chapters specifically devoted to
topics of wider interest, such as the bio-
logy of plasmids other than F, or bac-
teriophages. On the other hand, a careful
use of the index (thoughtfully included in
each volume) reveals that much of the
information that would be expected to be
found in such chapters is covered piece-
meal elsewhere.

Overall we and other colleagues are
very impressed by these books. As is usual
with endeavours of this kind the various

chapters vary in quality according to the
amount of effort and objectivity that their
authors have put into them, but the
majority are stimulating and readable yet
comprehensive. The editors have obvi-
ously put a lot of effort into this enterprise
and played an effective disciplinary role in
maintaining breadth without overlap, and
a uniform style. There is nothing com-
parable available on the market and
everyone working with, or teaching
about, *E.coli* and *S.typhimurium* will find
these books to be invaluable. Final year
undergraduates and postgraduate stu-
dents will also find them an excellent
resource. □

*B. M. Wilkins and R. H. Pritchard are in the
Department of Genetics, University of Leices-
ter, Leicester LE1 7RH, UK.*

In the blooming business

Jane Lewis

The Biology of Dinoflagellates. Edited by
F.J.R. Taylor. *Blackwell Scientific*: 1987.
Pp.785. £90, \$195.

WHEN passing through US immigration
controls to attend a meeting on toxic
dinoflagellate blooms, I was much sur-
prised to find that the officer interviewing
me knew something about the topic of the
conference. The gentleman's interest was
due to the potential health hazard of the
shellfish he much enjoyed. Toxic blooms
(or red tides) of dinoflagellates are just
one of the reasons this immensely diverse
group of the phytoplankton is the subject
of scientific and occasionally public atten-
tion. Other dinoflagellate features of in-
terest are their cysts, which are increas-
ingly used for the dating of sediments for
the oil industry, their spectacular symbio-
tic associations in coral reefs, their impres-
sive displays of bioluminescence and their
unusual cell biological features.

Two fields of research have been largely
responsible for the recent rapid growth in
the literature on dinoflagellates — the
study of blooms and the use of cysts in
stratigraphic dating. Both of these sub-
jects are multidisciplinary in nature, and
the appearance of a book on the general
biology of the group is thus especially
useful.

The Biology of Dinoflagellates covers 15
aspects of the subject written by 18
authors. The contributions divide into two
types, general biological reviews being
interspersed with accounts of more speci-
alist topics. Most facets of dinoflagellate
biology have a chapter devoted to them,
with the exception of evolution. Particu-
larly notable is the discussion of hetero-
trophic nutrition, which brings together

much scattered information on this neg-
lected subject; the phenomenon has long
been known in dinoflagellates, but only
recently has a clear mechanism been docu-
mented in certain species. The contribu-
tion on dinoflagellate behaviour is also
stimulating, as is that on the biochemistry
of the dinoflagellate nucleus. The final
chapter on cysts is a welcome inclusion,
giving a view of the group from the fossil
record.

Taxonomic problems are particularly
thorny when a group has been claimed by
both botanists and zoologists — where,
after all, does a motile group having both
photosynthetic and non-photosynthetic
members, many with cell walls, belong?
Throughout the book these difficulties are
dealt with clearly, and the editor (who
declares himself a "card-carrying protist-
ologist") has provided an appendix ex-
plaining recent changes to those less
familiar with taxonomic intricacies. Read-
ing the work as a whole there is some
repetition of material, but this is limited
and is sufficient to allow each chapter to
be studied on its own. The various contri-
butions are written with clarity, which
should allow researchers of any discipline
to get to grips with the subject matter,
and most of them provide food for thought
for the specialist, although there is some
unevenness in the depth of coverage.

I first saw mention of this book being
"in press" in 1983, and although it con-
tains some 2,500 references, the long
gestation period shows. Many of the
authors have managed to slip in the more
important recent references, but the book
is only reliably up to date until 1982.
Nevertheless this is a useful reference
work for researchers looking for a wider
understanding of dinoflagellates. I look
forward to another volume covering the
lost ground. □

*Jane Lewis, 70 Queens Road, Slough, Berkshire
SL1 3QP, UK, was formerly in the Department
of Biology, Royal Holloway and Bedford New
College, University of London.*

Radiation hazards on space missions

John R. Letaw, Rein Silberberg and C.H. Tsao

On the ambitious forays into space currently being contemplated, astronauts would encounter unfamiliar radiation environments. The risk of undue exposure would be diminished by spacecraft shielding.

THE US National Commission on Space has recently provided an exciting vision of the next half-century of space research and exploration¹. The long-term goals are the advancement of science, exploration and the eventual settlement of the Moon and Mars, and stimulation of space enterprise for the benefit of the people on Earth. Any such programme will require the presence of many men and women in space for long periods.

Radiation doses on space missions of long duration are very large in comparison with the allowed dose to the general public of 0.5 rem a year. Future astronauts will receive dose equivalents of up to 50 rem a year from galactic cosmic radiation (GCR), increasing long-term cancer incidence. Rare solar particle events could deliver lethal doses of up to 1,000 rem in one day. The harsh and unfamiliar radiation environments encountered in future programmes must therefore be understood and managed so as to lessen these health risks.

Space-station crew members will be exposed to radiation doses dominated by protons trapped in the Earth's radiation belts. In an orbit of inclination 28.5 degrees they are protected by the Earth's magnetic field from energetic particles emitted by the Sun during solar flares, and from GCR which permeates the galaxy. Future missions in polar orbits of higher inclination, in high-altitude geosynchronous orbits and outside the Earth's magnetic field to the Moon, Mars and libration points, will not enjoy this protection.

Apollo programme

These radiation risks were first identified² during the Apollo programme; however, the radiation environment was not well described and biological effects of heavy ions were poorly known. Light flashes in the eyes of Apollo astronauts³ caused concern when they were traced to interactions of heavy ions in the retina, vitreous humour and possibly the optic nerve. Although the radiation doses from these events were small, the magnitude of the biological effect had not been established.

The most intense energetic solar particle event ever recorded was in August 1972, within months of the Apollo 16 and

Apollo 17 missions. This event would have made it necessary to abort operations on the lunar surface. As with any such event, it could not have been anticipated in advance. It now serves as a baseline model for an anomalously large solar particle event.

In this article we present calculations of the radiation dose equivalents to astronauts from GCR and from energetic solar particle events. We extend previous results identifying GCR as a significant factor in the space radiation dose⁴. In particular, we have determined the components of the radiation dose and have calculated the relationship between dose and shielding thickness, proposing shielding requirements for future spaceflights. We recommend a storm shelter protected by at least 9 cm of aluminium (or its equivalent) on all spaceflights outside the magnetosphere. On long-duration flights, such as a Mars mission, all habitable spaces should be shielded with 7.5 cm aluminium (or its equivalent).

Galactic cosmic radiation consists of protons and the nuclei of heavier elements moving at velocities near the speed of light. Ultimately, these charged particles (or secondary products from their nuclear reactions) deposit energy in tissues by stripping electrons from water and organic molecules. Physical and chemical disruption may either kill or transform a cell, inducing cancer. If many cells are killed, radiation sickness may also result.

For the calculation of the effects of this radiation, three distinct procedures are required. First, the particle radiation environment must be known; we adopt the Naval Research Laboratory GCR model⁵, which gives the energy spectra of all elements at sunspot maximum and minimum (cosmic-ray minimum and maximum, respectively). Solar-flare environments are also adopted from that report. The August 1972 peak intensity is normalized to the time-integrated spectrum⁶ if a duration of 16 hours is assumed. The composite worst-case intensity⁷ includes high-energy protons which were observed in the February 1956 flare.

The second part of this calculation deals with transport of the cosmic-ray primary particles, and secondary particles produced in nuclear interactions, through shielding materials. A typical interaction

of a heavy ion in aluminium shielding material immediately liberates neutrons, protons and other light nuclei from the incident nuclei. Pions are often created in relativistic interactions. The residual fragments are in excited quantum states which decay by evaporation of light particles and emission of gamma radiation. Finally, long-lived radionuclides decay, usually by beta emission.

Transport codes

All components of radiation are transported using two nuclear transport codes (MCNP⁸ and HETC⁹) and procedures developed by us for heavy-ion transport^{10,4}. Our shielding model simulates the protection of the blood-forming organs (bone marrow) using a layered infinite plane slab geometry. The layers are: a variable thickness of aluminium shielding; 4.5 cm of water, approximating to body self-shielding; and a 1.0 cm water-absorption region. The density of energy deposited in the absorption region is proportional to the dose in rads.

The third part of this calculation is assessment of the biological effect of the dose by computing the dose equivalent in rem. This assessment depends both on the amount of energy deposited in the tissue and the relative biological effectiveness (RBE) of the radiation. RBE is the dose required by a given radiation to produce a specific biological effect, divided into the dose required by a reference radiation such as X-rays. The RBE of different particle species varies greatly, depending on the rate at which a particle deposits energy in tissue, the distribution of energy deposited in a cell, the rate at which the dose is received and the type of biological damage considered.

Heavy ions and neutrons have RBEs of 20 or more for many biological actions¹¹. Because cosmic-ray iron nuclei deposit energy in cells at 676 times the concentration that cosmic-ray protons do, a rare cosmic-ray iron nucleus is 13,500 times as dangerous as a cosmic-ray proton. This factor is comparable to the relative abundance of protons and iron nuclei in GCR. Heavy nuclei can therefore dominate the radiation dose equivalent to astronauts.

A great deal of research into the biological effects of heavy ions is under way,

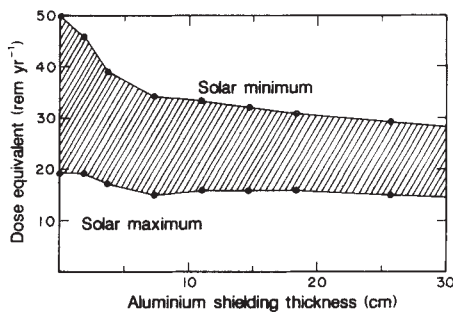


Fig. 1 Annual radiation dose equivalent from galactic cosmic radiation against aluminium shielding thickness. The shaded area illustrates the extremes of variation over the solar cycle. All secondary species are included in this estimate.

but RBEs for specific effects in human beings are often unknown¹². At present there is no universally accepted methodology for computing the overall health risk of a specified radiation exposure. For radiation protection applications an energy-dependent quality factor is assigned to each species. The quality factor effects the conversion from dose to dose-equivalent. Minimum ionizing radiations such as high-energy protons are assigned a quality factor of about one. In this work, the quality factors of primary heavy ions⁴ vary from 1 to 20, secondary protons are 1.2, and neutrons and nuclear recoils are 20. These figures are in good agreement with those proposed by the International Commission on Radiological Protection (ICRP)¹³.

Computed radiation doses to astronauts from galactic cosmic radiation are shown in Fig. 1. Annual dose equivalents with no aluminium shielding range from 20 to 50 rem, much greater than the 0.5 rem limit for the general population on Earth. Shielding 30 cm thick reduces the radiation dose by only 40 per cent at solar minimum and by less than 25 per cent at solar maximum. This result demonstrates that long-duration exposure to GCR will be a concern on many future space missions.

The US National Commission on Radiation Protection and Measurement (NCRP) has recommended an annual dose limit of 50 rem from all sources for astronauts on future missions¹⁴. The maximum GCR dose, coupled with other radiation exposures, exceeds the NCRP guidelines. There is also considerable uncertainty in the data and the procedures used to compute the risks of heavy ion exposure in space. We estimate that the uncertainty in our calculations is a factor of two. To assure conformity with the NCRP limit, we therefore recommend¹⁵ 7.5 cm of aluminium shielding (or its equivalent) for all habitable areas of any spacecraft designed for long-duration missions. An additional 20 cm of aluminium shielding would reduce the radiation dose by less than 20 per cent.

A cylindrical module 5 m in diameter and 9 m long (that is, about the size of the Space Shuttle bay) with the recommended aluminium shielding has a mass of about 33 metric tons. This shield could conceivably be lifted into space and propelled (say) to Mars, but only at great expense. Further consideration of shielding, better knowledge of space radiation and improved understanding of the biological effects of heavy ions may alter shielding requirements appreciably and possibly permit less severe measures. On the other hand, if large numbers of civilians with scientific or economic goals travel into space, dose limits more nearly in line with terrestrial limits must be adopted.

Figure 2 illustrates the components of the radiation dose at solar minimum. The character of the radiation changes within the shielding. Initially, the dose is dominated by heavy ions and protons; but after 10 cm of aluminium, secondary particles with large quality factors, including neutrons and nuclear recoils, dominate.

Various shielding strategies can be conceived which would limit the production of secondary particles and absorb neutrons. In particular, water (or methane) is a much better shield than aluminium because it yields fewer neutrons, is more effective at fragmenting and slowing down heavy ions, and is an efficient neutron moderator. Excellent thermal neutron absorbers, such as ¹⁰B, are well known. Considerations of this nature are crucial for reducing weight requirements in future spacecraft.

Using the quality factors adopted here, the radiation dose to unshielded astronauts during the August 1972 solar flare would have been 960 rem. (This value is consistent with, but higher than, other dose estimates¹⁶, because secondary production is considered.) With 9 cm aluminium shielding, the dose equivalent would have been 40 rem. The higher dose is lethal and radiation sickness would have prevented a mission from continuing¹⁷. The lower dose is survivable and, for most astronauts, would have resulted in no short-term health problems.

Anomalous large solar particle events occur about once a decade. For each week in space (outside the magnetosphere) there is consequently 1 chance in 500 that unshielded astronauts will receive a lethal radiation dose. We recommend that all manned spacecraft intended to spend periods of a week or more outside the magnetosphere should be equipped with a storm shelter providing 9 cm aluminium (or equivalent) shielding in all directions.

Our study of the (hypothetical) worst-case particle event finds the dose would be 94 rem per hour with no shielding and 22 rem per hour with 30 cm aluminium shielding. Exposure for 5 hours to this flare, even with the 9 cm thick storm

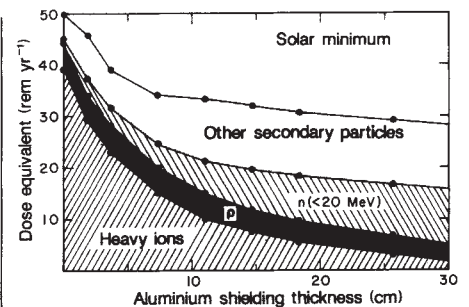


Fig. 2 Components of the cosmic radiation dose equivalent at solar minimum against aluminium shielding thickness. Doses from primary heavy ions, primary protons (p), secondary neutrons (energy < 20 MeV), and other secondary particles (including recoil nuclei and high-energy neutrons) are shown.

shelter recommended above, would severely disrupt a manned space mission. The particle fluxes during the February 1956 flare, on which the worst-case model is based, are uncertain by a factor of three. We do not base any recommendations on the worst-case model, but mention it to demonstrate the uncertainty about radiation hazards that may be encountered on future space missions.

We thank Paul Keaton, Michael Duke, Barney Roberts, D. Stuart Nachtwey and Harry Letaw, Jr for their comments and suggestions, and Scott Clearwater for his assistance in performing transport calculations. This work was supported by the Office of Naval Research and the Department of Energy. □

1. *Pioneering the Space Frontier* Rep. Natl Commission on Space (Bantam, Toronto, 1986).
2. Curtis, S.B. in *Space Radiation Biology* (eds Tobias, C.A. & Todd, P.) 21–99 (Academic, New York, 1974).
3. Pinsky, L.S. *et al. Science* **183**, 957–959 (1974).
4. Silberberg, R., Tsao, C.H., Adams, J.H. & Letaw, J.R. *Rad. Res.* **98**, 209–226 (1984).
5. Adams, J.H., Silberberg, R. & Tsao, C.H. *Cosmic Ray Effects on Microelectronics, Part I: The Near-Earth Particle Environment* Naval Research Laboratory Memorandum Rep. 4506 (Naval Research Laboratory, Washington, DC, 1981).
6. King, J.H. *J. Spacecraft* **11**, 401–408 (1974).
7. Adams, J.H. & Gelman, A. *IEEE Trans. Nucl. Sci.* **NS-31**, 1212–1216 (1984).
8. *MCNP (Monte Carlo Neutron and Photon Transport Code)* (Radiation Shielding Information Center, Oak Ridge National Laboratory, CCC-200, 1983).
9. Prael, R.E. *Users Guide to the HETC (Monte Carlo High Energy Nucleon-Meson Transport Code) Code System* Los Alamos Laboratory Draft Rep. (Los Alamos Laboratory, Los Alamos, New Mexico, 1985).
10. Letaw, J.R., Silberberg, R. & Tsao, C.H. *Astrophys. J. Suppl. Ser.* **56**, 369–391 (1984).
11. Sinclair, W.K. *J. Rad. Prot. Dos.* **13**, 319 (1985).
12. Todd, P. *Adv. Sp. Res.* **3**, 187–194 (1983).
13. *Recommendations of the ICRP* ICRP Rep. 26 (Pergamon, Oxford, 1977).
14. Fry, R.J.M. *Adv. Sp. Res.* **6**, 261–268 (1986).
15. Letaw, J.R. & Clearwater, S.H. *Radiation Shielding Requirements on Long-Duration Space Missions* Severn Communications Corp. Rep. 86–02 (Severn Communications Corp., Severn Park, 1986).
16. Atwell, W., Beever, E.R. & Hardy, A.C. *Proc. NATO Advanced Study Institute on Terrestrial Space Radiation and its Biological Effects* (Plenum, New York, in the press).
17. *Radiobiological Factors in Manned Space Flight* (ed. Langham, W.H.) 248 (Publ. 1487, National Academy of Sciences, Washington, DC, 1967).

John R. Letaw is at Severn Communications Corporation, PO Box 544, Severn Park, Maryland 21146, USA. Rein Silberberg and C.H. Tsao are at the E.O. Hulburt Center for Space Research, Naval Research Laboratory, Washington, DC 20375, USA.

Future emission scenarios for chemicals that may deplete stratospheric ozone

James K. Hammitt*, Frank Camm*, Peter S. Connell†, William E. Mooz*, Kathleen A. Wolf*, Donald J. Wuebbles† & Anil Bamezai*

* The RAND Corporation, 1700 Main Street, Santa Monica, California 90406-2138, USA

† Lawrence Livermore National Laboratory, PO Box 808, Livermore, California 94550, USA

Scenarios are developed for long-term future emissions of seven of the most important manmade chemicals that may deplete ozone and the corresponding effect on stratospheric ozone concentrations is calculated using a one-dimensional atmospheric model. The scenarios are based on detailed analysis of the markets for products that use these chemicals and span a central 90% probability interval for the chemicals' joint effect on calculated ozone abundance, assuming no additional regulations.

ENVIRONMENTALLY persistent halocarbons containing chlorine or bromine may reduce the concentration of stratospheric ozone¹⁻⁴ and contribute to global warming⁵. Industrially produced halocarbons that are believed to be among the most potent and are emitted in significant quantities include chlorofluorocarbons (CFCs) 11, 12 and 113, carbon tetrachloride, methyl chloroform and halons 1211 and 1301. Collectively, we call these substances 'potential ozone depleters' (PODs).

To date, most research on stratospheric ozone depletion has focused on reducing uncertainty about atmospheric photochemistry and physics. With a few exceptions^{4,6}, most of the reported research has assumed no future growth in POD emissions. This has been appropriate for testing and comparing atmospheric models, but as policymakers rely on these models, a more realistic treatment of emissions and related uncertainties is becoming increasingly important. Here we describe a way for projecting future POD emissions and report what we believe to be a reasonable range of future emission scenarios, assuming

no additional emission regulations. Such estimates can be used to gauge the importance of current uncertainty about future emission profiles in assessing the likely extent of future ozone depletion and compare it with other sources of uncertainty. In addition, they are a necessary first step in assessing whether additional regulations are appropriate. Technologies for reducing POD emissions and their costs have been analysed elsewhere⁷⁻¹⁰.

The emission scenarios are developed using the four-step process described below. They span a subjective 90% probability region for the model-calculated joint effect of the seven PODs on globally averaged stratospheric ozone abundance until 2040¹¹. (Unlike a confidence interval, a probability region describes an interval that one believes contains the unknown value of a variable with specified probability. It relies on bayesian as opposed to sampling theory.) The associated change in column ozone, as calculated using the Lawrence Livermore National Laboratory (LLNL) one-dimensional model, ranges between -1.8% and -17.3% in 2040. A subjective probability bound on future ozone depletion also requires analysis of uncertainty about other factors including the model specification, emissions of other trace gases and future social responses to perceived environmental risks.

PODs differ in the effectiveness with which a unit mass of each destroys ozone because of differences in molecular weight, the number of chlorine or bromine atoms per molecule, and the compounds' atmospheric lifetimes. The relative depletion efficiencies are estimated using the LLNL one-dimensional model. Table 1 illustrates one measure of their relative contributions to ozone depletion, expressed as the estimated share of current emissions weighted by the model-calculated depletion efficiency per unit mass. Other PODs (including CFC-22, 114 and 115) are quantitatively unimportant at current emission rates.

Individual chemical projections

Step 1. Initially we develop base projections and ranges of uncertainty for POD use in various applications and regions. These projections represent a consensus of the RAND authors. They are developed from preliminary scenarios for each market that consider trends in the product market and in POD use per unit product, as well as factors that historically have influenced them or may in the future influence them. Preliminary illustrating the range of reasonably foreseeable results were constructed and interpreted as bounding an 80% probability region. These rely on detailed information about specific POD applications obtained through literature review and contact with over 500 industry officials since 1978^{7-9,12-14}.

Table 1 Estimated relative contributions to ozone depletion

Chemical	Atmospheric lifetime (yr)*	Estimated global 1985 emissions (thousand tonnes)	Relative depletion efficiency*	Share of total contribution (%)
CFC-11 (CCl ₃ F)	76.5	238	1.00	25.8
CFC-12 (CCl ₂ F ₂)	138.8	412	1.00	44.7
Carbon tetrachloride (CCl ₄)	67.1	66	1.06	7.6
CFC-113 (CFC1 ₂ CF ₂ Cl)	91.7	138	0.78	11.7
Methyl chloroform (CH ₃ CCl ₃)	8.3	474	0.10	5.1
Halon 1301 (CF ₃ Br)	100.9	3	11.4	3.7
Halon 1211 (CF ₂ ClBr)	12.5	3	2.70	0.9
CFC-22 (CHF ₂ Cl)	22.0	72	0.05	0.4

Atmospheric lifetime = steady-state atmospheric burden/constant emission rate. Relative depletion efficiency = O₃ depletion per unit mass of species/O₃ depletion per unit mass of CFC-11. Share of total contribution = $100 \times S_i / \sum S_i$ where S_i = 1985 emissions $\times$ relative depletion efficiency. Shares do not sum to 100 due to rounding.

* Calculated using LLNL one-dimensional model.

Because of differences in available information, we distinguish between three world regions: the United States, the other countries whose CFC-11 and 12 production are reported by the Chemical Manufacturers Association¹⁵ (the other OECD nations and most developing countries) and the non-reporting countries (primarily the USSR, China and Eastern Europe). The three regions respectively account for an estimated 30%, 56% and 14% of current world production, weighted by relative depletion efficiencies. For the period before 2000, we develop separate projections and uncertainty ranges for each of the major POD applications in the first two regions and for total use of each POD in the non-reporting countries¹².

Such detailed analysis would become too speculative after 2000, so a more aggregate approach is used. Global use of most PODs is expected to grow with the global economy, and the uncertainty ranges reflect our uncertainty about the distant future relative to the period before 2000. The distribution for global economic growth is based on subjective probability distributions for global labour productivity and population developed by W. Nordhaus and G. Yohe (personal communication).

Subjective probability distribution

Step 2. Step 1 produces a range of potential growth for each chemical during the periods 1985–2000 and 2000–2040. In the second step, we partition this growth into two components corresponding to growth in general economic activity and chemical use relative to this growth, then we integrate across regions and product applications to obtain components describing the growth of gross global product (GGP) and global POD use relative to GGP. This allows us to account for the correlation between chemicals that results because each is correlated with general economic activity. The distribution of these components is described by a subjective multivariate normal probability distribution, the parameters of which are summarized in Table 2. We assume the growth rates for chemical use relative to economic growth are independent of one another and of general economic growth, although correlations could be integrated using this methodology.

Standard deviations are smaller after 2000 because the length of the second period reduces the variance of the mean growth rates relative to those in the first period despite higher uncertainty about the annual growth rates. The annual variance is 25% greater for each chemical after 2000 than before¹¹.

To understand better the considerations underlying the base projections, uncertainty ranges and corresponding probability distributions, consider the historical, economic, technological and regulatory factors associated with the use of each POD (for more detail, see ref. 12). Except for the sharp decline in aerosol use of CFC-11 and 12 (and in carbon tetrachloride used to produce these CFCs) that accompanied the ban on CFC use in aerosol products in the United States and other countries in the late 1970s, the available data suggest that growth in all seven PODs has been strong^{12,14,15}.

CFC-11 and 12 growth in foam manufacture and most other applications is expected to continue, but use in aerosol and some refrigeration applications should stagnate or decline. In part, because these chemicals have been in use for many decades, substantial new applications are not expected. Slower growth could result if CFC-11 and 12 are displaced from current applications (certain foam uses, for example). Higher growth could reflect longer market-expansion periods in the developing nations or the emergence of major new markets.

Carbon tetrachloride is used primarily to produce CFC-11 and 12. Solvent and fumigant uses have been discontinued in the United States because of concerns about health effects, but may continue elsewhere. Future use is linked to CFC-11 and 12 production.

CFC-113 production has expanded dramatically in recent years, largely because of its widespread use as a solvent in the electronics industry. The base projection assumes continuing

Table 2 Parameters of the subjective normal probability distribution for gross global product and chemical use relative to GGP

Chemical use relative to GGP	1985–2000		2000–2040	
	Mean	Standard deviation	Mean	Standard deviation
CFC-11	0.02	0.98	0	0.67
CFC-12	–1.00	1.05	0	0.72
Carbon tetrachloride	–0.49	1.02	0	0.70
CFC-113	3.27	1.67	0	1.15
Methyl chloroform	–0.32	1.10	0	0.75
Halon 1301	1.08	2.29	–0.45	1.57
Halon 1211	0.96	2.33	–0.05	1.60
Gross global product	3.28	1.15	2.40	0.96

Chemical growth rates are sum of growth rates of GGP and chemical use relative to GGP. Correlations among rates are zero. (Average growth rates in % per annum).

Table 3 Production growth rates for joint distribution scenarios

Chemical	Period	Quantile of the score function				
		0.05	0.25	0.50	0.75	0.95
CFC-11	Pre-2000	1.47	2.54	3.29	4.03	5.12
	Post-2000	0.93	1.80	2.40	3.00	3.86
CFC-12	Pre-2000	0.39	1.50	2.27	3.04	4.15
	Post-2000	0.89	1.78	2.40	3.02	3.91
Carbon tetrachloride	Pre-2000	0.93	2.03	2.79	3.55	4.64
	Post-2000	0.91	1.79	2.40	3.01	3.89
CFC-113	Pre-2000	4.15	5.56	6.55	7.54	8.97
	Post-2000	0.51	1.63	2.40	3.18	4.29
Methyl chloroform	Pre-2000	1.03	2.17	2.96	3.74	4.88
	Post-2000	0.86	1.77	2.40	3.03	3.94
Halon 1301	Pre-2000	1.12	2.94	4.07	5.34	7.15
	Post-2000	–0.20	1.07	2.00	2.92	4.22
Halon 1211	Pre-2000	1.12	2.94	4.07	5.34	7.15
	Post-2000	0.19	1.51	2.47	3.39	4.72

Growth rates in % per annum.

high growth, although the 1986 US ban on land disposal of chlorinated solvents could moderate US growth by increasing disposal costs. Higher growth could occur if competing solvents (such as trichloroethylene, perchloroethylene and methylene chloride) face additional regulations.

Methyl chloroform is used largely in solvent and adhesive applications. As for CFC-113, growth may be affected by substitution among solvents stimulated by the land-disposal ban or more stringent regulation of other chlorinated solvents. Because methyl chloroform is a mature chemical with well-established applications, growth should approximate that of GGP.

Halon 1301 has been rapidly adopted for use as a fire extinguishant in enclosed spaces such as computer rooms. Growth depends on the number and size of new systems and on whether the chemical is recovered and reused when systems are dismantled.

Halon 1211 is increasingly used in portable fire extinguishers. Recovery is unlikely because of the small quantities in individual units.

Joint production scenarios

Step 3. Production growth rates for individual chemicals are correlated with GGP growth, but not perfectly. Thus, it would be incorrect to combine production growth rates corresponding to the 95th percentiles of the distributions for each chemical and call this a '95th percentile' joint scenario: such an approach would lead to a larger model-calculated effect on ozone than is associated with the 95th percentile of our joint distribution.

Table 4 Production and (emissions) for joint distribution scenarios

Chemical and year	Quantile of the score function				
	0.05	0.25	0.50	0.75	0.95
CFC-11					
1985	342 (238)	342 (238)	342 (238)	342 (238)	342 (238)
2000	426 (309)	499 (358)	556 (396)	619 (438)	723 (507)
2040	618 (525)	1,017 (819)	1,435 (1,122)	2,022 (1,543)	3,294 (2,445)
CFC-12					
1985	444 (412)	444 (412)	444 (412)	444 (412)	444 (412)
2000	471 (467)	555 (541)	622 (599)	696 (662)	817 (765)
2040	672 (662)	1,125 (1,092)	1,606 (1,544)	2,287 (2,178)	3,788 (3,559)
Carbon tetrachloride					
1985	1,029 (66)	1,029 (66)	1,029 (66)	1,029 (66)	1,029 (66)
2000	1,183 (76)	1,391 (90)	1,554 (100)	1,736 (112)	2,032 (131)
2040	1,702 (110)	2,827 (182)	4,014 (259)	5,686 (366)	9,339 (601)
CFC-113					
1985	163 (138)	163 (138)	163 (138)	163 (138)	163 (138)
2000	300 (253)	367 (310)	423 (357)	485 (410)	591 (499)
2040	367 (310)	700 (591)	1,091 (922)	1,695 (1,432)	3,172 (2,680)
Methyl chloroform					
1985	545 (474)	545 (474)	545 (474)	545 (474)	545 (474)
2000	636 (554)	752 (654)	844 (734)	946 (823)	1,113 (969)
2040	898 (781)	1,517 (1,320)	2,179 (1,896)	3,123 (2,717)	5,215 (4,537)
Halon 1301					
1985	10.7 (3.0)	10.7 (3.0)	10.7 (3.0)	10.7 (3.0)	10.7 (3.0)
2000	13.6 (5.6)	16.7 (6.5)	19.4 (7.1)	22.4 (7.9)	27.7 (9.2)
2040	9.9 (8.5)	24.2 (14.7)	43.2 (22.0)	75.4 (33.6)	163 (62.9)
Halon 1211					
1985	10.7 (3.0)	10.7 (3.0)	10.7 (3.0)	10.7 (3.0)	10.7 (3.0)
2000	13.6 (5.6)	16.7 (6.5)	19.4 (7.1)	22.4 (7.9)	27.7 (9.2)
2040	17.5 (16.1)	34.1 (24.6)	55.1 (33.9)	89.6 (47.9)	182 (81.3)

Units are thousand tonnes.

The joint scenarios are derived from a 'score function' consisting of a weighted sum of the growth rates for each chemical. The weights reflect the compounds' capabilities to deplete ozone, based on each chemical's share of total production weighted by the fraction that is eventually emitted and its relative depletion efficiency. The weights change over time because of shifts in relative production, but begin with values similar to those in the last column of Table 1. (Because relative depletion efficiencies for the halons were not available, we used approximate efficiencies of 10.0 for both halons, 0.1 for methyl chloroform, and 1.0 for the other four PODs¹¹.)

The score function provides a rough measure of the potential effectiveness of these chemicals, taken together, in depleting stratospheric ozone. Moreover, we can derive a probability distribution for its value using the subjective probability distribution for individual chemical growth rates.

Scenarios are defined by quantiles of the score function's distribution. For example, the '95th percentile' scenario is constructed so the subjective probability that the value of the score function will be higher is 5%. Each value of the score function is associated with an infinite number of combinations of individual-chemical growth rates. We choose the unique set for which all of the GGP and chemical use components correspond to a common quantile of their respective marginal distributions. This choice distributes responsibility for the joint effect across all seven PODs: it associates relatively high- or low-growth scenarios for individual chemicals together.

Table 3 reports the production growth rates for various scenarios, and Table 4 presents the corresponding production levels. The procedure significantly narrows the range of scenarios, compared with an approach that does not adjust for joint variation. For example, the 95th percentile scenario uses growth rates drawn from the 82nd percentile of the distributions for individual chemicals. Analogously, the growth rates associated with low scenarios are raised.

Emission scenarios

Step 4. The timing of POD emissions varies by application. In some, PODs are released to the atmosphere shortly after manufacture. In others, they remain 'banked' in products until disposal. The last step introduces appropriate lags between production and emission for each POD. Resulting emissions are reported in Table 4.

CFC-11 and 12 applications can be divided into five categories. (1) Hermetically sealed applications (home refrigerators and freezers) have estimated typical lifetimes of 17 years during which small amounts of CFC-12 are emitted as individual units fail. (2) Non-hermetically sealed applications (retail-food refrigerators, building chillers, mobile air conditioners) lose CFCs through leakage and servicing throughout their lives, estimated to be seven years. (3) Urethane closed-cell foams emit CFC-11 very slowly. During their estimated 30-year service lives, emissions decline geometrically and are characterized by an assumed 100-year half life (based on measurements by Khalil and Rasmussen¹⁶). After disposal, emissions are assumed to accelerate to reflect the breakup of the foam into smaller pieces. Additional small amounts are emitted at manufacture. (4) Non-urethane foams release CFC-12 over an estimated two-year period. (5) Prompt emitters (aerosol propellants, flexible-foam blowing, sterilants) emit CFCs as they are used, shortly after manufacture.

Approximately 94% of carbon tetrachloride produced is used as an intermediate in the manufacture of CFCs. Most of the remainder is emitted promptly.

Most CFC-113 and methyl chloroform emissions are immediate, although an estimated 15% of production is lost in waste-disposal dumps, incinerated, or used as intermediates in manufacturing other chemicals.

Halons are largely banked in fire-extinguishing systems that have estimated service lives of up to 40 years. Small amounts

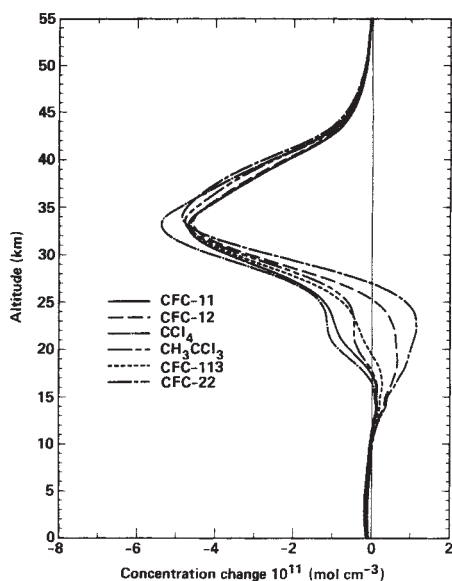


Fig. 1 Vertical profiles of calculated ozone concentration change for fixed emissions of six Cl-containing PODs at steady state. Surface POD emissions were normalized to produce the same abundance of inorganic Cl-containing species at 35 km in each case.

are emitted annually through leakage, fire, and inadvertent discharge, and at disposal.

Atmospheric model calculations

The joint emission scenarios provide boundary conditions for the LLNL one-dimensional model of transport, photochemical kinetics and radiative processes in the troposphere and stratosphere. The model is representative of those used for calculating future ozone changes and has been described previously^{6,17,18}. Its domain extends from the Earth's surface to 55 km, and it includes 165 photolytic and thermal reactions of 52 chemical species, a diffusive representation of globally averaged transport, and interactive calculation of the temperature profile above 14 km (assuming radiative equilibrium). All significant known homogeneous photochemistry of the oxygen, nitrogen oxide, chlorine, bromine and methane families is included. Kinetic rate constants and photochemical parameters are based on the recommendations of the NASA Panel for Data Evaluation¹⁹. From an initial condition representing the current (1985) atmosphere, integration over time produces the model-predicted global annual average changes in tropospheric and stratospheric ozone abundances. We also use the model to calculate relative depletion efficiencies for the seven PODs to compare with those used in developing the joint scenarios.

The calculated change in total column ozone represents an annual global average of the seasonally and latitudinally varying changes that are expected to result from POD and other trace gas emissions. One-dimensional model results are useful for delineating the ranges of calculated ozone change that arise from uncertainties about future POD emissions and other factors, although higher-dimensional models are preferred for estimating effects on the biosphere.

Model-calculated relative depletion efficiencies

Relative ozone-depletion efficiencies play an important role in constructing the score function above. They can be calculated using the one-dimensional model by comparing total column ozone change calculated at steady state for a specified emission of a particular POD to that calculated for any reference POD; we used CFC-11. At steady state, the surface emission and atmospheric destruction of the POD are balanced, and an equilibrium vertical profile of the POD abundance is established. The model-calculated atmospheric lifetime is equal to the ratio of

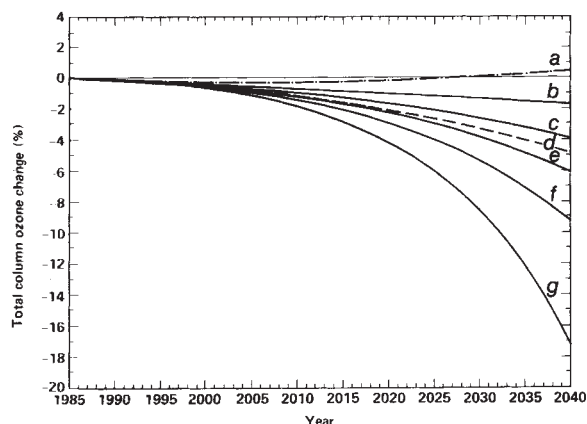


Fig. 2 Calculated one-dimensional total column ozone change for seven POD emission scenarios. Assumed growth rates for other trace gases are consistent with current trends: CO₂, 0.64% per yr; N₂O, 0.25% per yr; CH₄, 1% per yr; CFC-22, 0% per yr. *a*, Constant future POD emissions; *b*, 0.05 quantile of score function; *c*, 0.25 quantile of score function; *d*, 0.50 quantile of score function with CFC-11, 12 production cap; *e*, 0.50 quantile of score function; *f*, 0.75 quantile of score function; *g*, 0.95 quantile of score function.

the steady-state atmospheric burden to the prescribed emission rate.

In determining relative efficiencies, emission levels were chosen to produce similar calculated total column ozone depletions for each species to minimize differences caused by other effects such as the 'self-healing' increase in lower stratospheric ozone when upper stratospheric ozone is depleted. Specified surface emissions for each POD were integrated over a period of about three POD-dependent atmospheric lifetimes. The calculated relative efficiencies by emitted mass and atmospheric lifetimes are reported in Table 1.

Because the active agent in POD-caused ozone depletion is the chlorine or bromine released in the stratosphere, relative efficiencies by mass depend on the parent molecule's molecular weight, number of chlorine and bromine atoms, and atmospheric lifetime. These factors account for most of the variation in relative efficiency. Figure 1 shows that the ozone-depletion profiles for CFC emissions normalized for differences in these factors are roughly equivalent. Some of the remaining spread can be related to the vertical profiles of atomic chlorine release by photolysis of the CFCs. CFC-22 and 12 appear slightly less efficient in decreasing the total column ozone on this normalized basis because they have higher average altitudes for photolysis, and thus contribute less chlorine to the lower stratosphere than the other compounds. Smaller chlorine abundances in the lower stratosphere are currently thought to increase ozone abundances locally as a result of interactions with the dominant ozone-loss processes, the reactions of the nitrogen oxides.

Model-calculated ozone depletion

To calculate the predicted effects on ozone of the joint POD emission scenarios, trends for other atmospheric trace species must be assumed. The ambient 1985 atmosphere is obtained using industrial emission estimates for the CFCs and observed trends for CO₂, N₂O and CH₄ abundances as boundary conditions¹⁷. For future growth in trace gas abundances we use the CO₂ projection in Edmonds *et al.*²⁰ (averaging 0.64% per yr) and assume continuation of currently estimated global average trends in N₂O (0.25% per yr) and CH₄ (1% per yr)¹⁸. Surface emissions of CFC-22 are fixed at the current rate of ~71.9 thousand tonnes per yr. The effects of uncertainties in these assumptions on calculated ozone change are discussed by Connell and Wuebbles¹⁷.

Figure 2 shows changes in model-calculated total column

Table 5 Calculated surface abundances for joint emission scenarios

Chemical	Year	Zero emission growth	Quantile of the score function					0.50 with production cap
			0.05	0.25	0.50	0.75	0.95	
			Parts per billion					
CFC-11	1985	0.206	0.206	0.206	0.206	0.206	0.206	0.206
	2000	0.300	0.332	0.346	0.356	0.367	0.385	0.356
	2040	0.508	0.770	0.995	1.21	1.48	2.01	0.860
CFC-12	1985	0.377	0.377	0.377	0.377	0.377	0.377	0.377
	2000	0.626	0.653	0.676	0.694	0.713	0.742	0.694
	2040	1.18	1.46	1.87	2.27	2.77	3.77	1.66
Carbon tetrachloride	1985	0.153	0.153	0.153	0.153	0.153	0.153	0.153
	2000	0.158	0.161	0.164	0.167	0.169	0.174	0.167
	2040	0.166	0.199	0.249	0.297	0.358	0.480	0.297
CFC-113	1985	0.033	0.033	0.033	0.033	0.033	0.033	0.033
	2000	0.089	0.113	0.124	0.132	0.141	0.156	0.132
	2040	0.199	0.364	0.538	0.718	0.967	1.51	0.718
Methyl chloroform	1985	0.138	0.138	0.138	0.138	0.138	0.138	0.138
	2000	0.175	0.191	0.211	0.226	0.242	0.268	0.226
	2040	0.201	0.303	0.478	0.637	0.861	1.31	0.641
Parts per trillion								
Halon 1301	1985	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2000	3.39	4.99	5.43	5.75	6.13	6.73	5.75
	2040	10.3	22.8	31.8	40.6	52.9	80.6	40.6
Halon 1211	1985	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2000	2.10	3.23	3.55	3.78	4.06	4.50	3.78
	2040	3.15	14.9	19.8	24.8	31.8	46.8	24.9
Other trace gases (all scenarios) parts per million								
	Year	CO ₂	CH ₄	N ₂ O				
	1985	345	1.74	0.303				
	2000	372	2.03	0.315				
	2040	490	3.02	0.348				

Quantities are given in moles of chemical per mole of atmosphere.

ozone for the joint emission scenarios. Also shown are the calculated changes resulting from constant POD emissions at 1985 rates, and from the 50th percentile scenario with a cap on CFC-11 and 12 production at the estimated current world capacity, 1.24 million tonnes²¹. Table 5 reports the corresponding calculated surface abundance of the seven PODs and other trace gases at selected dates.

The range of calculated ozone changes corresponding to the 5th and 95th percentile scenarios spans an order of magnitude (−1.8% to −17.3%) by 2040. At −6.2%, the 50th percentile scenario falls nearer the geometric than the arithmetic mean of the other scenarios, as the distribution of ozone depletion is skewed to the smaller depletion side. This skewness reflects the skewness of emission levels, resulting from the symmetry of growth rates across scenarios.

Even if POD emissions are held to 1985 levels, POD abundances continue to increase because of the long atmospheric lifetimes. The initial small decrease in column ozone abundance in this case is followed by a continuing small increase as CO₂ and CH₄ abundances increase. Increasing CO₂ cools the stratosphere and slows ozone-loss processes. Increasing CH₄ has the net effect of increasing ozone as a by-product of methane oxidation in the lower stratosphere and troposphere, and by sequestering active chlorine species in the inactive form of HCl.

Capping CFC-11 and 12 production at estimated current capacity in the 50th percentile scenario allows production of these species to increase until 2003, after which they remain constant. Ozone depletion continues, reaching −4.9% in 2040, about 1.3 percentage points less than under the uncapped scenario. Stratospheric inorganic chlorine abundances continue to increase because of increasing emissions from the bank, the CFCs' long atmospheric lifetimes, and increasing emissions of other PODs.

The range of calculated ozone depletion represents a subjective 90% probability interval for the calculated joint effect of the seven PODs on stratospheric ozone, assuming no additional regulations. It represents that part of our uncertainty about future depletion associated with current uncertainty about future uncontrolled POD emissions. A probability interval for future ozone depletion requires assessment of other factors that contribute to that uncertainty, including trends in abundances of other gases, the photochemical and kinetic processes modelled, and possible regulations.

Conclusions

Uncertainty about future POD emissions is substantial. But because the compounds' growth rates are not perfectly correlated, uncertainty about their calculated joint effect on ozone is proportionally smaller than uncertainty about the calculated effect of any one of them. Moreover, it appears highly likely that this joint effect will continue to grow, assuming no additional regulations. The commonly used zero-emission-growth scenario falls below the 5th percentile of the distribution of the score function; even though emissions of some PODs may decline, we estimate a greater than 95% chance that total weighted emissions will increase.

A significant factor underlying this result is that almost all observers expect general economic growth to continue over the period analysed¹⁴. Our low-growth scenarios combine lower-than-expected economic growth with declining POD use relative to GGP. These low average rates are reasonable to expect only if major changes in regulation, product innovation, or other forces offset the general trend of global economic growth.

The emission levels shown in Table 4 suggest that CFC-11 and 12 will continue to be the chemicals most likely to contribute

to any ozone depletion that occurs, as they are at present. Such a conclusion, however, requires further analysis. The method used to construct scenarios groups low growth rates for individual chemicals together into low scenarios and high rates into high scenarios; this creates more stable shares than can reasonably be expected. Because the analysis focuses on total ozone depletion and not on the exact PODs that are responsible, this approach is appropriate. But it does not address the relative importance of each chemical.

Received 6 April; accepted 12 October 1987.

1. Molina, M. J. & Rowland, F. S. *Nature* **249**, 810–812 (1974).
2. Cicerone, R. J., Stolarski, R. S. & Walters, S. *Science* **185**, 1165–1167 (1974).
3. Stolarski, R. S. *Am. Stat.* **36**, 303–311 (1982).
4. Prather, M. J., McElroy, M. B. & Wofsy, S. C. *Nature* **312**, 227–231 (1984).
5. Dickinson, R. E. & Cicerone, R. J. *Nature* **319**, 109–115 (1986).
6. Wuebbles, D. J. *J. geophys. Res.* **88**, 1433–1443 (1983).
7. Palmer, A. R., Mooz, W. E., Quinn, T. H. & Wolf, K. A. *Economic Implications of Regulating Chlorofluorocarbon Emissions from Nonaerosol Applications* R-2524-EPA (The Rand Corporation, Santa Monica, 1980).
8. Mooz, W. E. *et al. Technical Options for Reducing Chlorofluorocarbon Emissions* R-2879-EPA (The Rand Corporation, Santa Monica, 1982).
9. Camm, F. *et al. Social Cost of Technical Control Options to Reduce Emissions of Potential Ozone Depleters in the United States: An Update* N-2440-EPA (The Rand Corporation, Santa Monica, 1986).
10. Hammitt, J. K. *Timing Regulations to Prevent Stratospheric-Ozone Depletion* R-3495-JMO/RC (The Rand Corporation, Santa Monica, 1987).
11. Camm, F. & Hammitt, J. K. *An Analytic Method for Constructing Scenarios from a Subjective Joint Probability Distribution* N-2442-EPA (The Rand Corporation, Santa Monica, 1986).
12. Hammitt, J. K. *et al. Product Uses and Market Trends for Potential Ozone-Depleting Substances, 1985–2000* R-3386-EPA (The Rand Corporation, Santa Monica, 1986).
13. Kohler, D. F., Haaga, J. & Camm, F. *Projections of Consumption of Products Using Chlorofluorocarbons in Developing Countries* N-2458-EPA (The Rand Corporation, Santa Monica, 1987).
14. Quinn, T. H. *et al. Projected Use, Emissions and Banks of Potential Ozone Depleting Substances* N-2282-EPA (The Rand Corporation, Santa Monica, 1986).
15. Chemical Manufacturers Association *Production, Sales and Calculated Release of CFC-11 and CFC-12 through 1985* (Chemical Manufacturers Association, Washington DC, 1986).
16. Khalil, M. A. K. & Rasmussen, R. A. *J. Air Pollut. Control. Ass.* **36**, 159–163 (1986).
17. Connell, P. S. & Wuebbles, D. J. *Ozone Perturbations in the LLNL One-Dimensional Model—Calculated Effects of Projected Trends in CFCs, CH₄, CO₂, N₂O, and Halons over 90 Years* UCRL-95548 (Lawrence Livermore National Laboratory, Livermore, 1986).
18. World Meteorological Organization *Atmospheric Ozone 1985: Assessment of Our Understanding of the Processes Controlling its Present Distribution and Change* Report No. 16 (Global Ozone Research and Monitoring Project, Washington DC, 1986).
19. DeMore, W. B. (ed.) *Chemical Kinetics and Photochemical Data for Use in Stratospheric Modeling* Publication 85-37 (Jet Propulsion Laboratory, Pasadena, 1985).
20. Edmonds, J. A., Reilly, J., Trabalka, J. R. & Reichle, D. E. *An Analysis of Possible Future Atmospheric Retention of Fossil Fuel CO₂* Technical Report TR013 (US Department of Energy Carbon Dioxide Research Division, Washington DC, 1984).
21. Mooz, W. E., Wolf, K. A. & Camm, F. *Potential Constraints on Cumulative Global Production of Chlorofluorocarbons* R-3400-EPA (The Rand Corporation, Santa Monica, 1986).
22. Crawford, M. *Science* **237**, 1557 (1987).

Structure and evolution of the northern Red Sea

R. W. Girdler & T. C. Southren

School of Physics, The University, Newcastle upon Tyne NE1 7RU, UK

Combined analyses of new geophysical data in the northern Red Sea with data from the Gulf of Suez and the Gulf of Aqaba–Dead Sea rift strongly suggest that the development of the Red Sea took place in three stages: (1) a ‘Gulf of Suez’ stage in the Oligocene, accompanied by thinning of the continental lithosphere and the formation of an early ‘Red Sea–Gulf of Suez’ graben; (2) an early ‘Dead Sea’ stage, with slow seafloor spreading from ~25 to 15 Myr accompanied by 62 km of shear along the Aqaba–Dead Sea transform; and (3) a recent ‘Dead Sea’ stage, with slow seafloor spreading from 4.5 to 0 Myr accompanied by 45 km of shear along the Aqaba–Dead Sea transform.

THE nature of the crust beneath the Red Sea, especially in the north, remains the subject of considerable controversy. Some suggest that it is mostly underlain by oceanic crust^{1–3}; others maintain that most of it is stretched and thinned continent^{4–9}. The situation is complicated by the very large thicknesses of evaporites which affect the magnetic signature of the underlying rocks¹⁰. To help resolve this question, 44 new bathymetric, gravity and magnetic profiles were obtained between the Sinai peninsula and 19° N in 1979 aboard RRS *Shackleton* (see Fig. 1 inset). Navigation was controlled by satellite fixes and radar fixes on the coasts and islands. All the profiles are in a north-easterly direction and extend as near to the coasts as possible. Here, some of these are combined with neighbouring land data from Africa and Arabia to construct long profiles which are used in an attempt to determine the structure of the Red Sea, especially in the north.

New contour maps¹¹ of bathymetry, gravity and magnetic anomalies reveal that the northern Red Sea can be subdivided at 24° and 25° N into three regions (Fig. 1). Region 1, in the north, has a relatively simple main trough structure; region 2 between 24° and 25° N, is a complex transform zone offsetting regions 1 and 3 by ~100 km; and region 3 to the south has an axial trough within the main trough⁴.

Region 1. Here the coasts are straight, with the bathymetry

showing two distinct trends: parallel to the coasts near the margins, and oblique to the coasts in the central region. Near the coasts, the water depth increases rapidly to 500 m, with the contours parallel to the coasts. In the central region, a line of small deeps extends transversely across the Red Sea from the mouth of the Gulf of Suez to 25.5° N, trending N39°W (as compared with N28°W for the west coast and N32°W for the east coast), that is, approximately at right angles to the Aqaba–Dead Sea transform. The deep water extends close to the Sinai peninsula, the 1,000-m contour being within 5 km of the Sinai coast. The Hume Deep is parallel to the deeps of the Gulf of Aqaba, suggesting that the Aqaba structures continue into the northernmost Red Sea. The free-air and Bouguer gravity anomalies also show the two trends, but the magnetic anomalies show only the oblique trend. The Bouguer anomalies are ~0 to +20 mGal near the coasts and increase to +60 mGal over the deep water, with several superimposed short-wavelength (15–20 km) anomalies of +20 to +40 mGal amplitude. The magnetic field is characterized mainly by long-wavelength, low-amplitude anomalies plus several small-wavelength, large-amplitude anomalies reaching 1,300 nT.

Region 2. Here all the features become complex, the bathymetry having an approximate NNE trend. The free-air, Bouguer and magnetic anomalies are all very subdued. The region is located

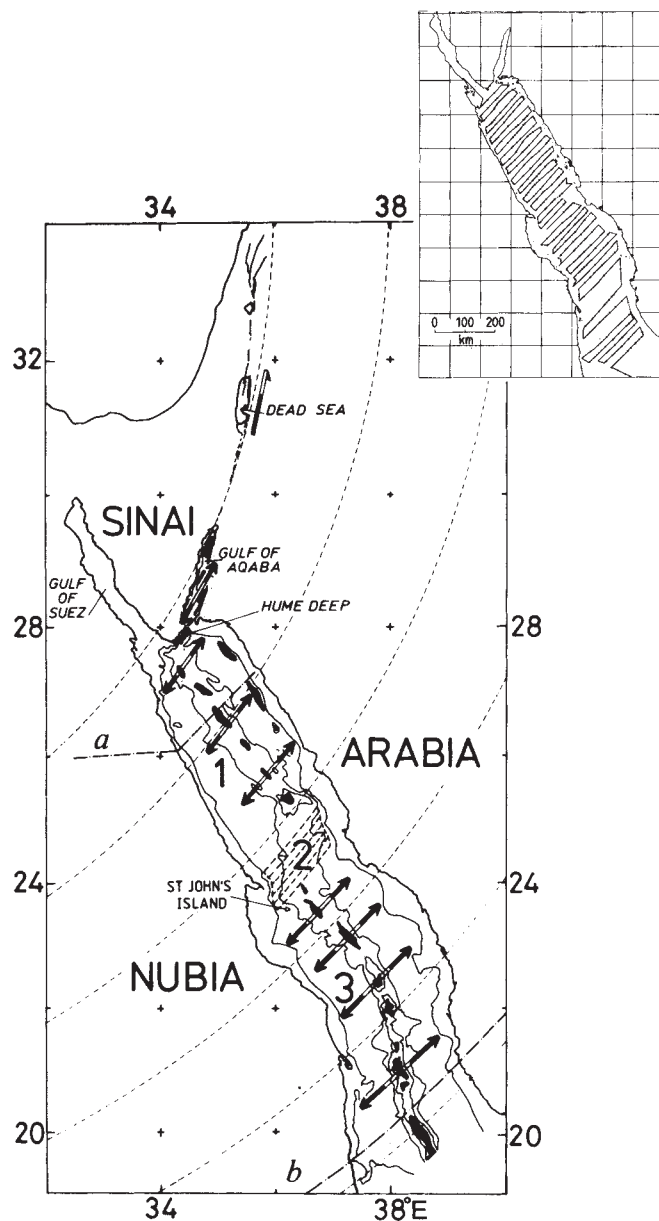


Fig. 1 Map of the northern Red Sea, showing the location of the complex transform fault zone (region 2) offsetting regions 1 and 3 by 80–100 km. The dashed lines are small circles about the Dead Sea pole of rotation¹⁸ at 33° N, 24° E, which describes the motion of Arabia with respect to Sinai. The Hume Deep in the northernmost Red Sea and the individual deeps in the Gulf of Aqaba are all parallel to the small circles, whereas the deeps in the central Red Sea tend to be perpendicular to the small circles. The lengths of the arrows correspond to the two stages of movement (62 and 45 km) along the Dead Sea transform. Black segments correspond to the first state, white to the second. Dot-dashed lines mark the locations of the gravity profiles shown in Fig. 2. Inset, chart of ship's track along which bathymetric, gravity and magnetic profiles were collected.

where there is an 80–100 km right-lateral offset of the Red Sea and is clearly a complex transform zone extending over 50–60 km.

Region 3. At 24.5° N, the coasts become sinuous and all the features become parallel to them (in contrast to region 1). The axial trough becomes prominent and widens towards the south. There are a number of deeps along the axis which trend perpendicular to the direction of separation of Arabia from Africa (see Fig. 1). The Bouguer anomalies are generally positive, reaching

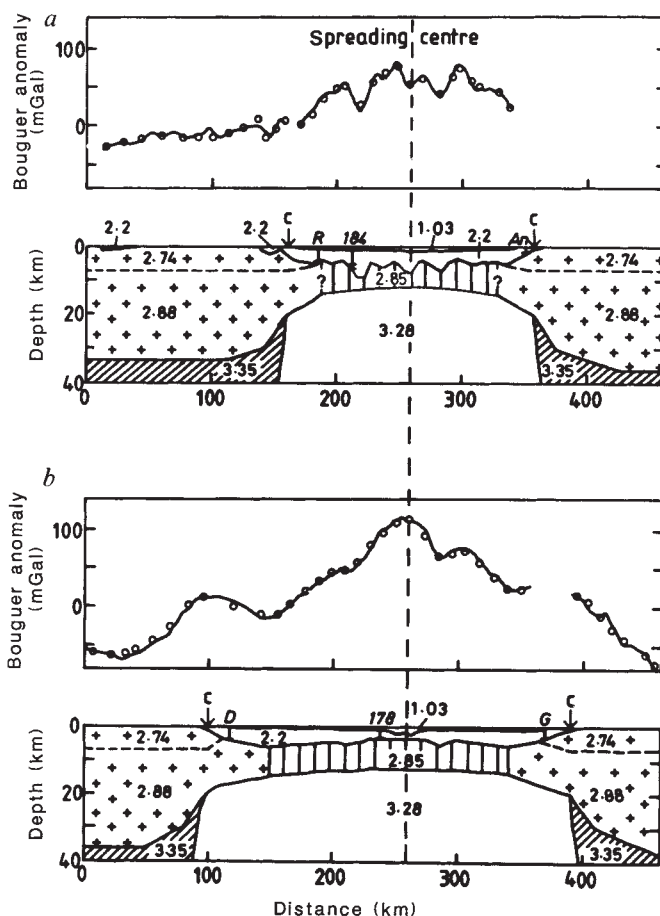


Fig. 2 Bouguer gravity anomaly profiles over regions 1(a) and 3(b) with possible interpretations consistent with the plate tectonic setting shown in Fig. 1. C denotes the positions of the coasts, other letters and numbers the locations of boreholes and seismic refraction profiles, respectively. Crosses, continental crust; vertical lines, oceanic crust; diagonal lines, uppermost mantle; white, asthenosphere with densities in g cm⁻³. The locations of the profiles are shown in Fig. 1 as dot-dashed lines.

a maximum of +120 to +140 mGal over the axial trough. On either side of the maximum there are often minima ~20–50 km from the coasts. The magnetic anomalies over the main and axial troughs are much larger (300–1,500 nT), and many profiles show symmetry about the axial anomaly (Fig. 3). This suggests the presence of seafloor spreading, implying that a large proportion of the crust in this region is oceanic.

St John's Island, with its +164 mGal Bouguer anomaly¹², is in the north-west corner of the region (Fig. 1). It seems generally agreed that the St John's peridotite is a post-Mesozoic fragment of the mantle^{12–15}, as originally proposed by Moon¹⁶. Recent radiometric measurements¹⁵ suggest that its age is late Oligocene or Miocene. The neighbouring gneisses are considered to be either fragments of continental basement^{13,14} or metamorphosed sediments and volcanics above the mantle diapir¹⁵. There are two views concerning its tectonic setting, with some^{12–15} considering it to be associated with a fracture zone, and others¹⁵ considering that it was emplaced in a rift setting. The latter is in agreement with its location in region 3 (Fig. 1) although the major fracture zone (region 2) is close by.

Regional setting

In the north, the Red Sea bifurcates into the Gulfs of Suez and Aqaba, with the Sinai peninsula in between (Fig. 1). Any model

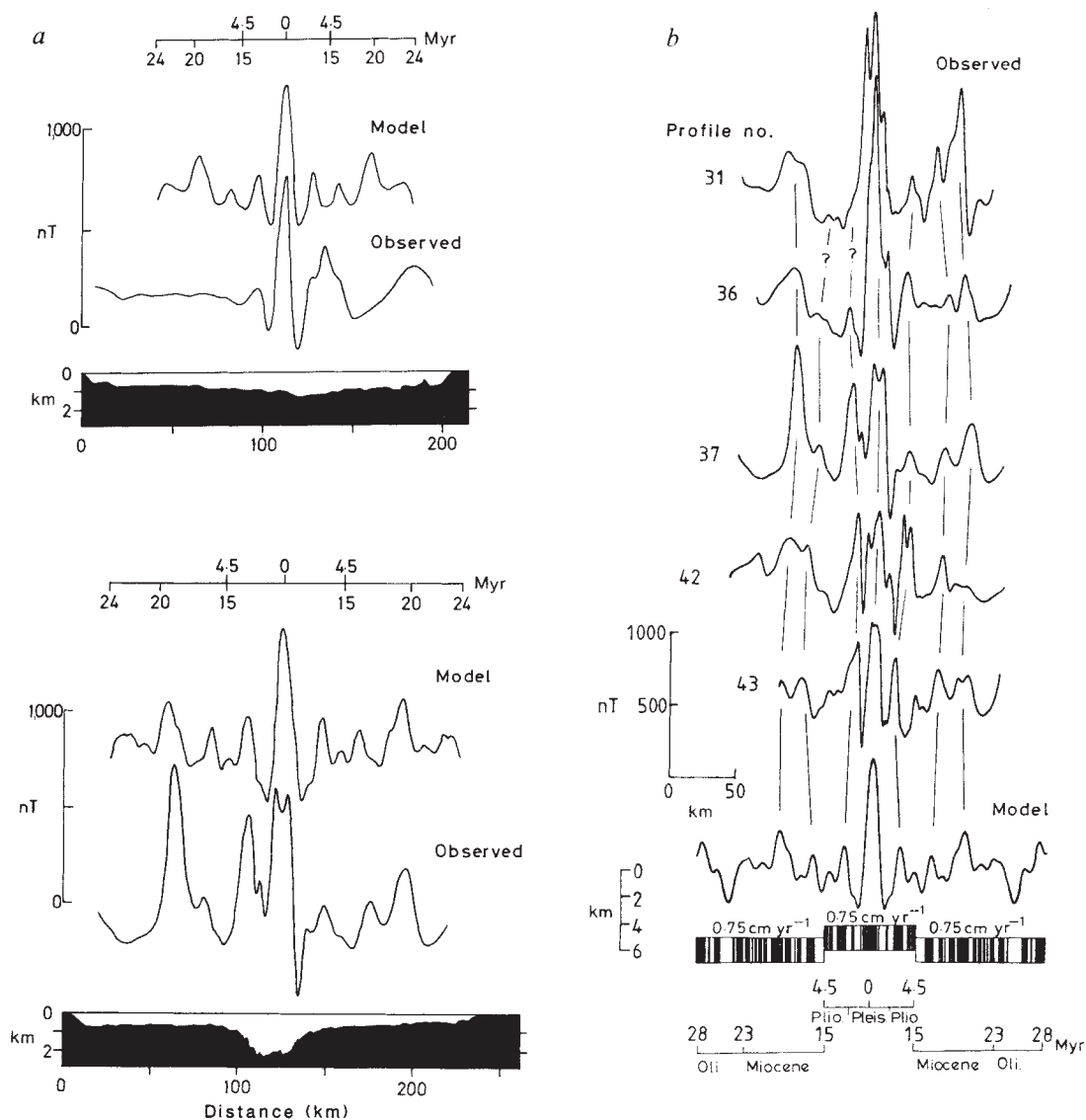


Fig. 3 Examples of total-intensity magnetic anomaly profiles, converted to the pole. *a*, Relationships to bathymetry for region 1 (above) and region 2 (below); *b*, correlations of the anomalies over region 3. In both *a* and *b*, seafloor spreading models consistent with the two-stage movement along the Dead Sea transform (with a hiatus between 15 and 4.5 Myr) are shown for comparison.

for the structure and evolution of the Red Sea must therefore be consistent with the structure and evolution of these two regions.

The western marginal faults of the Red Sea and Gulf of Suez (Clysmic) rift are in continuity, and therefore must have been formed at the same time. The Clysmic Gulf is ~80 km wide and bounded by normal faults. Rift movements began some time after the end of the Eocene (38 Myr) and continued through the Oligocene, leading to a major unconformity at the base of the Miocene (25 Myr)¹⁷. There are several dykes and sills, the dykes being mostly parallel to the main trend of the graben. These intrude the Mesozoic and Eocene and terminate abruptly at the base of the Miocene; they must therefore be Oligocene in age. During the Miocene, the Gulf became infilled with sediments, including large thicknesses of evaporites¹⁸; the present water depth is <60 m. The loading contributed to subsidence, which was particularly rapid from 25 to 15 Myr, with further subsidence from 5 to 0 Myr in the south¹⁹. Since the Lower Miocene²⁰ there has been no igneous activity and the Gulf appears to have become less active.

In contrast, the Gulf of Aqaba makes an angle of 40° to the western margin of the Red Sea and is seismically active. It forms part of the Dead Sea transform zone between the opening Red Sea to the south and the collision zone of the Taurus-Zagros mountains to the north. The zone consists of a number of rhombochasmms formed as Arabia moved north with respect to

Sinai²¹. Geological mapping and radiometric dating²² suggest that this movement has occurred since latest Oligocene/early Miocene in two stages, with a quiet period in between. The early movement (62 km) is difficult to date but is considered to be earliest Miocene²³; the second movement (45 km) is Plio-Pleistocene and still continuing²⁴. Quennell²⁵ noticed that the arcuate faults all have "a common centre of rotation at 33° N, 24° E". This describes the motion of Arabia with respect to Sinai. If Arabia is restored by the total movement of 107 km, the western coast of Arabia aligns with the eastern coast of the Gulf of Suez and the Precambrian marginal scarps align perfectly²⁶. The pole of Quennell²⁵ must therefore describe the opening of the central Red Sea after the formation of the Gulf of Suez. The small circles for this pole are shown in Fig. 1 (dashed lines), along with the corresponding amounts of opening to be expected at various latitudes in the Red Sea (arrows). It can be seen that the deeps in the Gulf of Aqaba and Hume Deep in the northernmost Red Sea are all parallel to the small circles, whereas the deeps in the centre of the Red Sea tend to be at right angles to them. The transform zone at 24–25° N is also parallel to the small circles.

Gravity profiles

Gravity profiles over regions 1 and 3 have been modelled using the above and borehole and seismic refraction data as constraints. Of the 38 boreholes reported in the Red Sea^{27–30} only

13 have reached basement, and these are all north of 19° N. North of 24° N (Fig. 5), four boreholes within 25 km of the Egyptian coast and four boreholes within 25 km of the Arabian coast reached Precambrian basement at depths between 2.2 and 5 km. One borehole, 55 km off the Egyptian coast at $\sim 25.8^{\circ}$ N, bottomed in oceanic gabbro. It would appear that in this area the transition between continental and oceanic crust lies between 25 and 55 km from the coast. Seismic refraction studies on land^{31,32} suggest that the continental crust thins rapidly beneath the coastal plains towards the Red Sea. At sea, seismic refraction experiments have shown that high-velocity oceanic basement is present beneath the main and axial troughs^{4,33}. A detailed seismic refraction experiment³³ which covered an approximately 1 degree square around 22.5° N, 33° E showed oceanic crust continuous beneath the axial and main troughs to within 65 km of the Arabian coast; a recent seismic refraction experiment in the northern Red Sea also showed the presence of oceanic crust³⁴. Taken together, these suggest that the crust beneath the coastal regions is of thinned continental type, whereas oceanic crust is present beneath the central regions, including the northern parts.

In addition to the well known positive Bouguer gravity anomaly over the Red Sea, the gravity profiles south of 26° N show a gravity minimum which occurs between 20 and 50 km from the coast, a feature also noticed by Izzeldin for a survey between Sudan and Arabia³⁵. From the above constraints it would appear that this feature is present at about the position where the transition between continental and oceanic crust might occur; that is, it might indicate the boundary between down-faulted continental crust and oceanic crust. Gravity interpretations are obtained consistent with these observations (see Fig. 2).

Figure 2 shows two examples of the gravity profiles, one for region 1(a) and one for region 3(b). Both show the major positive Bouguer anomaly, but over region 1, several smaller anomalies are superimposed on this. For the models, the major positive anomaly is assumed to be associated with the denser intrusive zone (producing oceanic crust), and the small anomalies with irregularities at the crust/sediment interface, the rough topography probably being related to slow spreading. The structure of the lithosphere away from the region of rifting is assumed to be the same as that used by Brown and Girdler³⁶ in their interpretation of African gravity.

The profile for region 3 (Fig. 2b) is at 20° N (see Fig. 1 for location). The rise in Bouguer anomaly towards the Red Sea was modelled as the thinning of the continental lithosphere, with the rate and amount of crustal thinning based on the seismic work of Makris *et al.*³¹ at 26° N. Seaward of the coasts the Bouguer anomaly decreases to a minimum ~ 50 km from the coasts, before rising steadily to a maximum of 100 mGal over the axial trough. The gravity minimum is interpreted as the boundary between thinned, down-faulted continental basement and oceanic basement, and the model predicts ~ 190 km of oceanic crust at 20° N measured along azimuth N44°. For the profile over region 1 (Fig. 1 and Fig. 2a) the main positive Bouguer anomaly is interpreted in an analogous way to that over region 3 and with the positions of the ocean-continent boundary consistent with the 107-km shear along the Dead Sea transform. The profile provides a good example of the higher-frequency anomalies superimposed on the long-wavelength positive anomaly. These tend to increase in amplitude and number to the north, being more prominent in region 1 than in region 3. Here these are interpreted as due to relief on the oceanic basement such as is found over slower-spreading ridges³⁷.

Although only two profiles are shown, the interpretation can be applied to many profiles within region 3 and some in the southern part of region 1. The positions of the ocean-continent boundary have been located for as many profiles as possible and are shown in Fig. 4a.

For the profiles over region 1 other interpretations are poss-

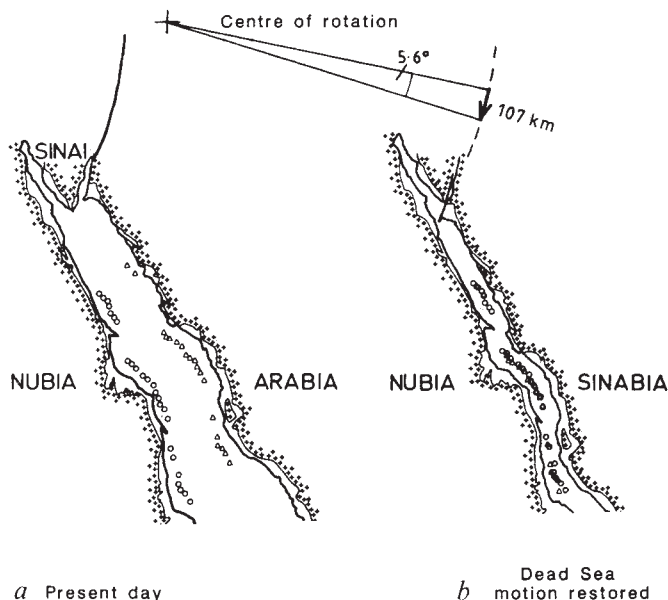


Fig. 4 Comparisons of the positions of the ocean-continent boundaries for Africa (O) and Arabia (Δ) estimated from the gravity models for individual profiles (a) before and (b) after restoration of Arabia with respect to Sinai by 107 km along the Dead Sea transform. Crosses (+) represent outcropping Precambrian basement and Mesozoic sediments.

ible, such as the presence of very thin, block-faulted continental crust. In this case, the smaller density contrast between the sediments and (lighter) sialic basement implies much greater relief. A further possibility is that the seafloor spreading in region 3 gives way to a continent-ocean mix (Precambrian basement plus recent intrusions) in region 1 in the north. The simplest interpretation, shown in Fig. 2a, is preferred, as it is consistent with the remarkable alignment of the Precambrian of the eastern Gulf of Suez and western Arabia on restoration of the 107-km Dead Sea shear movement.

Magnetic profiles

Figure 3 shows a selection of total-intensity magnetic anomaly profiles. The relationships of the anomalies to the bathymetry for regions 1 and 3 are shown in Fig. 3a and the way in which the anomalies can be correlated from profile to profile for region 3 in Fig. 3b. The symmetry and correlations strongly suggest the presence of seafloor spreading.

Because the Red Sea is narrow, the profiles are necessarily short, making it difficult to match the anomalies with synthetic seafloor spreading models. For this reason the models rely heavily on the geological boundary conditions from the Suez and Dead Sea rifts. As seen above, these suggest a three-stage evolution for the Red Sea, with (1) an early Gulf of Suez graben stage, mainly in the Oligocene; (2) opening, corresponding to the first 62 km of shear along the Dead Sea rift in the latest Oligocene and early Miocene; and (3) further opening, corresponding to the recent 45-km shear along the Dead Sea rift in the Plio-Pleistocene. The seafloor spreading model for the profiles over region 3 uses these constraints. As can be seen in Fig. 3b, reasonable fits to the observations can be obtained using the last two stages, that is, 4.5 to 0 Myr and 23 to 15 Myr.

In general, the magnetic anomalies over region 1 do not show good lineations, but if the synthetic seafloor spreading sequence developed for region 3 is used (Fig. 3a), the observed and synthetic anomalies compare reasonably well for the recent phase of spreading; beyond this, the correlations are poor. From this it could be argued that the oceanic crust is confined to pockets near the axis, but this is in conflict with the regional geological data. A possible explanation for this lies in the

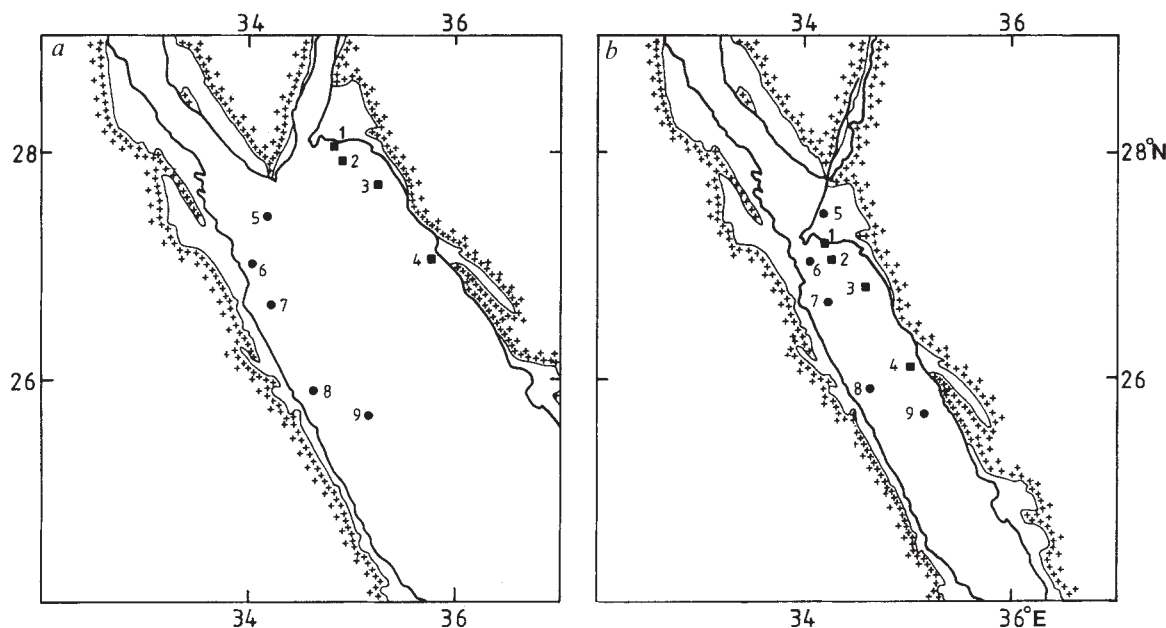


Fig. 5 The locations of boreholes in, *a*, the present-day northern Red Sea and, *b*, the Red Sea with the 170-km left-lateral shear along the Dead Sea transform restored. Crosses as on Fig. 4. Note the remarkable collinearity of the Precambrian for the western margins of Sinai and Arabia (south of Sinai). Borehole details²⁷⁻³⁰ (number, name, total depth (m), bottom-hole rock type): 1, Al Kurmah, 3,122, granite; 2, Barquan, 2,971, granite; 3, Yuda-1, 2,271, granite; 4, Annuman-1, 1,868, schist; 5, RSO T'95-1, 1,699, dacite; 6, RSO X'94-1, 2,847, granite; 7, RSO B96-1, 4,212, granodiorite; 8, A-IX, 5,040, schist; 9, B-IX, 4,180, gabbro.

presence of large thicknesses of unstable evaporites coupled with the slow spreading rates expected in the north. Although evaporite deposition ceased at the end of the Miocene (5 Myr), evaporites of Miocene age were found overlying oceanic crust with magnetic anomaly ages of <2.5 Myr at DSDP sites 227 and 228, indicating that the salt must have flowed³⁸. The salt flowage could have a profound effect on the formation of magnetic anomalies if the flow rate exceeds the seafloor spreading rate¹⁰. In this case, the spreading centre will be covered by evaporites, and the new oceanic crust will intrude into them, cooling slowly and giving a low remanent magnetization (in contrast to the usual quench-cooled basalts which give a large magnetic signature). This situation probably occurs in region 1, where the expected spreading rate is only $0.45\text{--}0.65\text{ cm yr}^{-1}$. Support for this comes from the relatively smooth bathymetry in region 1, in contrast to the rugged bathymetry of the axial trough in region 3, with its thin veneer of sediments. The presence of large anomalies over the deeps even in the most northerly part also support this interpretation.

Plate kinematic reconstruction

With the amount of separation estimated from the gravity anomalies and timings from the magnetic anomalies, a plate kinematic reconstruction can be attempted. As the magnetic anomalies can be interpreted to indicate seafloor spreading in the Red Sea consistent with the timings of major shear movements along the Dead Sea rift, and the formation of the Gulf of Suez is known to pre-date these movements, it is appropriate first to restore Arabia with respect to Sinai. In Fig. 4, Arabia is rotated clockwise by 5.6° (equivalent to the total 107-km shear along the Dead Sea rift). When this is done the positions of the African and Arabian ocean-continent boundaries estimated from the gravity interpretations fit remarkably well. This further supports the view that the motion of Arabia with respect to Sinai forming the Dead Sea transform was accompanied by the formation of oceanic crust in the Red Sea. Although the ocean-continent boundary could not be determined north of 26.5°N , boreholes that reach basement there show no conflict with this

interpretation. In Fig. 5, these boreholes are plotted with respect to the coasts for the present Red Sea (*a*) and for the reconstruction (*b*). As can be seen, there is no overlap between boreholes bottoming in Precambrian rocks. The "oceanic" gabbro found in borehole 9, 55 km off the Egyptian coast³⁰, now appears nearest to the Arabian coast. It follows that its formation was associated with the rotation of Arabia with respect to Sinai and the opening of the Red Sea, consistent with its oceanic character.

It remains to calculate the pre-rift configuration of Nubia, Arabia and Sinai. The remarkable alignment of the Precambrian margins of western Sinai and Arabia after the total movement along the Dead Sea rift has been completely restored (Fig. 5*b*) suggests that there was once a single continental plate consisting of Sinai and Arabia, which we will call Sinabia. The evolution of the Suez rift and the early Red Sea is thus a consequence of the movement of Sinabia with respect to Nubia (Africa). As the interpretation of the gravity data suggests that at this stage the northern Red Sea was entirely underlain by thinned continental lithosphere, the amount of extension can be estimated by reconstituting the thinned continental lithosphere in the gravity models to its pre-rift thickness. If this is done an average extension of 60 km is calculated, a factor of two greater than the maximum estimated extension across the Suez rift³⁹. The estimate is considered to be a maximum, as lithospheric thinning may have proceeded with asthenosphere actively replacing thinned continental lithosphere. It is difficult to estimate a pole of rotation for this early movement, but the direction of motion is assumed to be that which would produce the best fit of the coastlines. If the Sinabian and Nubian coastlines are moved together by 60 km, no overlap of the exposed Precambrian basement occurs and on average the coastlines are ~ 25 km apart (Fig. 6*a*). This represents the pre-rift configuration of Nubia, Sinai and Arabia but it should be stressed that it is strongly dependent on the value of lithospheric extension estimated from the gravity models. A pre-rift configuration based on the estimates of extension across the Suez rift is also possible, but note that the amount of extension in the Red Sea should be greater

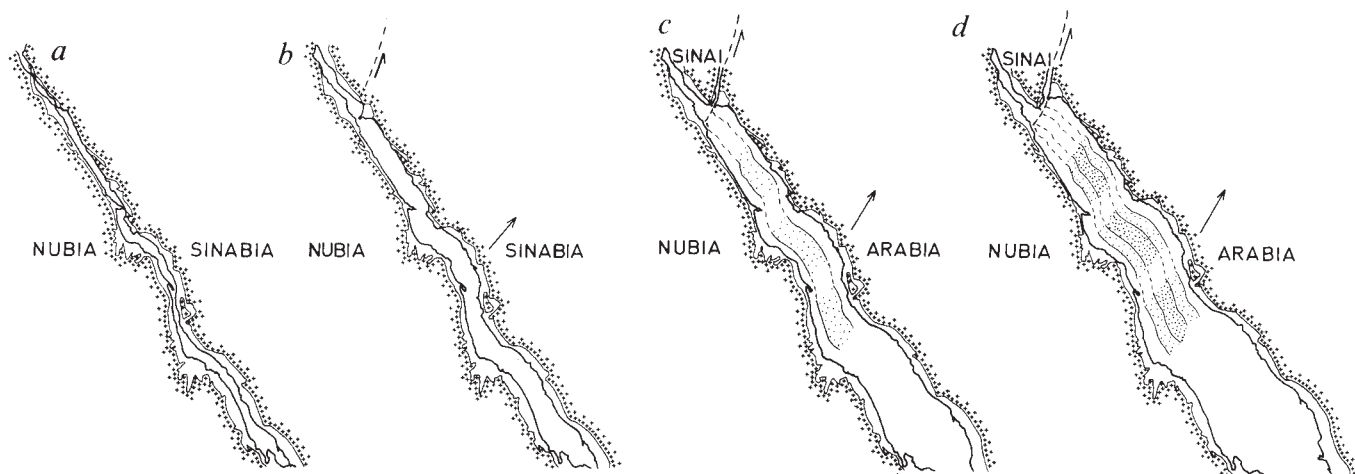


Fig. 6 The evolution of the Red Sea, based on marine geophysical data from the Red Sea and regional geological data, especially from the Suez and Aqaba-Dead Sea rifts. *a*, Early 'Gulf of Suez-Red Sea rift valley' stage in the Oligocene, ~30 Myr ago. *b*, The extended Red Sea with mainly stretched and thinned continental lithosphere at the beginning of the Miocene, ~25 Myr ago; cracks propagate NNE to form the early Gulf of Aqaba-Dead Sea rift. *c*, After the first major episode of seafloor spreading (shown with stipple) and 62 km of shear along the Dead Sea transform, ~15 Myr ago. *d*, The present Red Sea after the second phase of seafloor spreading (closer stipple), which was initiated at the end of the Miocene ~5 Myr ago, together with a further 45 km of shear along the Dead Sea transform.

because the Suez-early Red Sea rift widens towards the south, consistent with the rotation of Sinabia with respect to Nubia.

Evolution of the northern Red Sea

The structure of the Red Sea north of 19° N can be adequately explained by a three-stage evolution: an early period of continental extension followed by two periods of seafloor spreading (see Fig. 6). The pre-rift reconstruction of Nubia and Sinabia suggests that the present coastlines were ~25 km apart (Fig. 6*a*). Rifting began in the Oligocene (possibly late Eocene), forming the early Red Sea and Suez rifts. The continental lithosphere thinned in response to rifting and by the time the Suez rift reached its maximum width the early Red Sea rift was ~80–140 km in width (Fig. 6*b*).

It would appear that the onset of shear along the Dead Sea rift in the latest Oligocene/early Miocene was accompanied by the ending of major extension in the Suez rift. The component of shear introduced by this new direction of motion seems to have resulted in the complete fracture of Nubia and Sinabia, with the formation of oceanic crust—that is, the first stage of seafloor spreading from 23 to 15 Myr (Fig. 6*c*). This was followed by a spreading hiatus of ~10 Myr, during which the Red Sea was an enclosed basin supplied by influxes of water from the Mediterranean, allowing a great accumulation of evaporites.

Seafloor spreading started again ~4.5 Myr ago, the new axis of spreading coinciding approximately with the old axis, splitting the earlier oceanic crust into two (Fig. 6*d*). This event also marked the end of evaporite deposition in the Red Sea, owing to the influx of waters from the Gulf of Aden through the Straits of Bab el Mandeb. Although evaporite deposition ceased, evaporite flowage continued with the new phase of spreading. North of 23–24° N the flowage has kept pace with seafloor spreading, resulting in almost complete coverage of the oceanic crust, whereas to the south, where the seafloor spreading rates are faster, an axial trough 25–50 km in width developed mainly free of evaporites.

Recent analyses⁴⁰ of SIR-A Space Shuttle Radar and Landsat imagery confirm earlier studies⁴¹ stressing the importance of lineaments parallel to the Red Sea and roughly parallel to the Dead Sea trend in the neighbouring Precambrian. It seems likely that the early Red Sea rift (stages *a* and *b* in Fig. 6) may have been influenced by the N300–340° lineaments and the subsequent stages (*c* and *d*) by the N0–20° lineaments.

We thank Dr Paul Morgan of Northern Arizona University, Flagstaff for kindly providing land gravity data, Dr Douglas Robson for helpful discussions on the Gulf of Suez and Dr Enrico Bonatti for helpful comments on the manuscript. This work was supported by the UK NERC.

Received 19 June; accepted 23 September 1987.

- McKenzie, D. P., Davies, D. & Molnar, P. *Nature* **226**, 243–248 (1970).
- Hall, S. A., Andreasen, G. E. & Girdler, R. W. *Direct. Gen. Miner. Resour. Bull.* **22**, F1–F15 (1977).
- Styles, P. & Hall, S. A. in *Geodynamic Evolution of the Afro-Arabian Rift System* Accademia Nazionale Dei Lincei, Rome, 47, 587–606 (1980).
- Drake, C. L. & Girdler, R. W. *Geophys. J. R. astr. Soc.* **8**, 473–495 (1964).
- Ross, D. A. & Schlee, J. *J. Geol. Soc. Am. Bull.* **84**, 3827–3843 (1973).
- Lowell, J. D. & Genik, G. J. *Am. Ass. Petrol. Geol. Bull.* **56**, 247–259 (1972).
- Cochran, J. R. *Bull. Am. Ass. Petrol. Geol.* **67**, 41–69 (1983).
- Bonatti, E. *Nature* **316**, 33–37 (1985).
- Pautot, G., Guennoc, P., Coutelle, A. & Lyberis, N. *Nature* **310**, 133–136 (1984).
- Girdler, R. W. *Tectonophysics* **116**, 95–108 (1985).
- Southren, T. C. thesis, Univ. Newcastle upon Tyne (1986).
- Styles, P. & Gerdes, K. D. *Earth planet. Sci. Lett.* **65**, 353–368 (1983).
- Bonatti, E., Hamlyn, P. & Ottonello, G. *Geology* **9**, 474–479 (1981).
- Bonatti, E., Ottonello, G. & Hamlyn, P. *J. geophys. Res.* **91**, 599–631 (1986).
- Nicolas, A. & Boudier, F. *J. geophys. Res.* **92**, 461–474 (1987).
- Moon, F. W. *Preliminary Geological Report on St. John's Island (Red Sea)* (Geological Survey of Egypt, Government Press, Cairo, 1923).
- Robson, D. A. *Jl geol. Soc. Lond.* **127**, 247–276 (1971).
- Heybrook, F. in *Salt Basins Around Africa* 17–40 (Institute of Petroleum, London, 1965).
- Moretti, I. & Colletta, B. *J. Geodynam.* **7**, 151–168 (1987).
- Steen, G. *Egyptian General Petroleum Co. 6th Explo Seminar*, Cairo (1982).

- Quennell, A. M. *Q. Jl geol. Soc. Lond.* **114**, 1–24 (1958).
- Bartov, Y., Steinitz, G., Eyal, M. & Eyal, Y. *Nature* **285**, 220–222 (1980).
- Quennell, A. M. in *The Geology of Jordan* (eds Abed, A. M. & Khaled, H. M.) 460–482 (Jordanian Geologists Association, Amman, 1983).
- Garfunkel, Z., Zak, I. & Freund, R. *Tectonophysics* **80**, 1–26 (1980).
- Quennell, A. M. *Int. Geol. Congr. 20th session, Mexico*, 1956, 385–405 (1959).
- Girdler, R. W. in *The Geology of Jordan* (eds Abed, A. M. & Khaled, H. M.) 503–522 (Jordanian Geologists Association, Amman, 1983).
- Coleman, R. G. *Geological Map of the Red Sea* (US Geological Survey, Reston, 1973).
- Barakat, H. & Miller, P. M. *Egyptian General Petroleum Company, Seventh exploration seminar*, Cairo (1984).
- Khattab, M. M. *Mar. Geol.* **52**, 187–199 (1983).
- Noy, D. J. M. thesis, Univ. Newcastle upon Tyne (1979).
- Makris, J., Allam, A. & Moller, L. *Eos* **62**, 230 (1981).
- Mooney, W. D., Gettings, M. E., Blank, H. R. & Healy, J. H. *Tectonophysics* **111**, 173–246 (1985).
- Tramontini, C. & Davies, D. *Geophys. J. R. astr. Soc.* **17**, 225–241 (1969).
- Gaulier, J. M. *et al. Eos* **67**, 1208–1209 (1986).
- Izzeldin, A. Y. thesis, Univ. Louis Pasteur, Strasbourg (1982).
- Brown, C. & Girdler, R. W. *J. geophys. Res.* **85**, 6443–6445 (1980).
- Hall, S. A., Casey, J. F. & Elthorn, D. L. *J. geophys. Res.* **91**, 3724–3738 (1986).
- Girdler, R. W. & Whitmarsh, R. B. *Init. Rep. DSDP* **23**, 913–921 (1974).
- Robson, D. A. *Nature* **228**, 1237 (1970).
- Crane, K. & Bonatti, E. *J. geol. Soc. Lond.* **144**, 407–420 (1987).
- Dixon, T. H. & Stern, R. J. *J. Afr. Earth Sci.* (1984).

T-cell receptor δ gene rearrangements in early thymocytes

Yueh-hsiu Chien*, Makio Iwashima*, Daniel A. Wettstein*, Kenneth B. Kaplan*, John F. Elliott*, Willi Born† & Mark M. Davis*‡

* Department of Medical Microbiology and ‡ Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California 94305-5402, USA

† Howard Hughes Medical Institute, National Jewish Center for Immunology and Respiratory Medicine, Denver, Colorado 80206, USA

The T-cell receptor δ -chain variable region can be assembled from as many as four distinct gene segments, V, D₁, D₂ and J, more than any other antigen-receptor gene. In fetal thymocytes V→D joinings are as common as D→J or VDJ rearrangements and one V gene segment predominates. Analysis of rearrangements at TCR γ and δ loci during fetal ontogeny suggests abrupt changes and possible coordinate control in the rearrangement and expression of these loci.

RECENTLY we have identified a new T-cell receptor (TCR) gene termed 'X' (ref. 1). Like other TCR and immunoglobulin genes, it is assembled by somatic rearrangements which bring together variable (V) and joining (J) segments adjacent to a constant (C) region. This locus begins 75 kilobases (kb) 5' to C α and is in the same transcriptional orientation as the J α s and C α . That the X locus is entirely contained within the α gene is supported by the preliminary studies which indicate that the TCR X gene can use at least one V gene segment from a previously identified TCR α -gene family¹. Rearrangement and expression of this region occurs at least as early as day 14 in fetal thymocytes, well before any C α messenger RNA or protein is detected¹ and this locus is deleted in most α : β bearing cells². Both protein sequencing³ and analysis with anti-peptide antisera^{4,5} have now shown that X encodes the δ chain of the γ : δ T-cell receptor. This type of heterodimer appears early in T-lymphocyte differentiation and on a small percentage (3–5%) of normal peripheral T cells^{6–12}. Whereas α : β heterodimers have been found on all functional helper and cytotoxic T cells¹³ and appear necessary and sufficient for the recognition of antigens in association with major histocompatibility complex (MHC) products^{14,15}, the function of the γ : δ T-cell receptor is not clear. To understand better the assembly and repertoire of the δ -chain gene, we have analysed rearranged δ genes from both day-18 and day-19 fetal thymocytes as a population, and

also those within individual cells using a panel of day-14 to day-18 fetal thymocyte hybridomas.

Two D elements 5' of J_{δ1}

Analysis of δ -locus rearrangements (see next section) revealed two distinct D-region elements (Fig. 1). D_{δ1} is 9.8 kb and D_{δ2} is 0.9 kb 5' of the previously identified J_{δ1} element. D_{δ1} is 11 nucleotides and D_{δ2} is 16 nucleotides long. Both can be read in all three reading frames and are flanked by sequences related to the conserved heptamer and nonamer of immunoglobulins^{13,16}. The spacing between the heptamer and nonamer is 12 nucleotides on the 5' side and 23 nucleotides on the 3' side. Thus, the joining of the two D regions, as well as either or both of the D regions to J_{δ1} (which has a spacing of 13 nucleotides at the 5' side), would be consistent with the 12/23 rule of antigen-receptor joining^{17,18}. That both of these D regions can be used in a VDJ joining event is evident in analysis of the p λ -12 complementary DNA clone¹. At least 14 nucleotides can be accounted for as of D_{δ1} or D_{δ2} origin as indicated in Fig. 1c. A one-nucleotide gap was introduced to maximize the contribution of D_{δ2}, a difference which may be due to a strain polymorphism between the p λ -12 clone (B10.A) and the BALB/c derived germline sequences. Many other VDJ_δ cDNA clones have also been obtained¹⁹. To date no other antigen receptor has been shown to be the product of this many

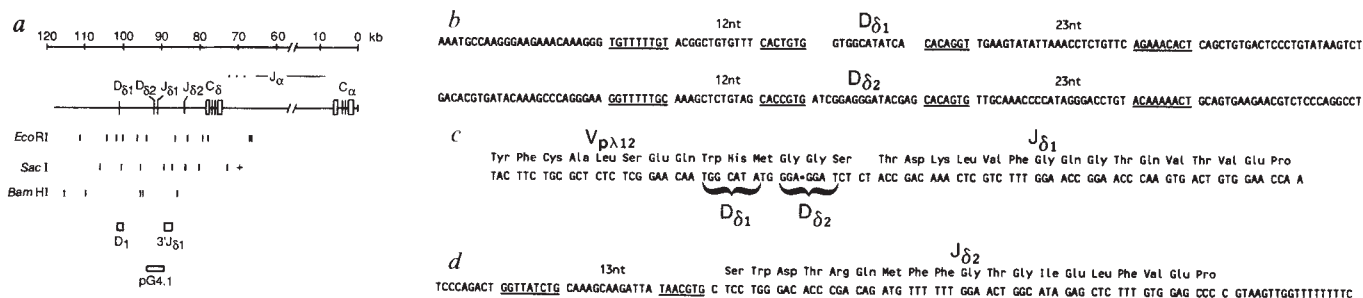


Fig. 1 Genomic organization of the TCR δ locus, sequences of the D and J_{δ2} regions and location of probes. *a*, A restriction map of the δ locus, its location with respect to J α - and C α -coding regions and the identification of the J_{δ1}- and C_δ-coding regions has been described previously¹. The positions of D_{δ1}, D_{δ2} and J_{δ2} are determined by restriction enzyme mapping, Southern analysis and DNA sequencing. The exon structure of C_δ is from Iwashima *et al.* (manuscript submitted). The DNA fragments used for further analysis of the rearrangements are: D_{δ1}, a 1.9 kb EcoRI subclone containing D_{δ1}; pG4.1, a 4.1 kb EcoRI to SacI subclone containing D_{δ2} and J_{δ1}; 3'J_{δ1}, a 2 kb SacI to SacI subclone located just 3' of J_{δ1}. These positions are indicated by open boxes. *b* and *d*, The sequences of D_{δ1}, D_{δ2}, J_{δ2} with heptamer, nonamer and RNA splice sequences underlined. All the DNA fragments were subcloned into the pGEM-3 vector (Promega Biotec), and sequenced by the dideoxy chain-termination method with both T7 and Sp6 promoter-specific primers on double-stranded plasmid DNA⁴⁵. All the sequences reported were done for both strands. To facilitate sequencing, deletional subclones were generated by ExoIII and mung-bean nuclease treatment⁴⁶. *c*, Alignment of the sequence between the V and J regions of a p λ -12 cDNA clone¹ with that of D_{δ1} and D_{δ2}. A gap was introduced and indicated by a dot to maximize the homology.

independently recombining gene segments and only two others have as many as three (IgH and TCR β). It is also interesting to note that all three types of antigen-receptor pairings (heavy and light chains of immunoglobulin, α : β and γ : δ of the T-cell antigen receptors) combine D-region containing polypeptides with non-D species, possibly reflecting some use for antigen recognition in such juxtapositions.

Identification and rearrangements of $J_{\delta 2}$

A new J_{δ} element ($J_{\delta 2}$) was identified through analysis of a C_{δ} -containing cDNA clone isolated from an adult double-negative thymocyte cDNA library¹⁹. It is located 7.2 kb 3' to the previously identified $J_{\delta 1}$ and 5.6 kb 5' to C_{δ} (Fig. 1a). The sequence of $J_{\delta 2}$ is presented in Fig. 1d. Other than the conserved amino acids of TCR and immunoglobulin J regions, it has very little similarity with $J_{\delta 1}$ or any of the published J_{δ} s (ref. 13). Conserved heptamer and nonamer sequences separated by 13 nucleotides are located 5' of the unrearranged $J_{\delta 2}$ sequence. A conserved RNA splicing-signal sequence is located at the 3' boundary. The last nucleotide of $J_{\delta 2}$ (c) differs from that of $J_{\delta 1}$ (a) resulting in different amino acids at the JC junction (glutamine versus lysine respectively), similar to the TCR α chain¹³.

Rearrangements at the $J_{\delta 2}$ locus are no more than one-tenth as frequent as those detected at $J_{\delta 1}$ as determined by Southern blots of fetal thymocyte DNAs (data not shown). But joining of a particular V region (V_{M21}) (see next section) and $J_{\delta 2}$ is observed in fetal thymocyte DNAs (data not shown) and in several γ : δ -bearing T-cell lines (J. Allison and R. O'Brien, personal communication).

Rearrangements at $J_{\delta 1}$

Rearrangements detected by a $J_{\delta 1}$ -containing fragment appear at least as early as day 14 in fetal thymocytes, and the pattern of bands becomes more intense and complex thereafter¹. A bacteriophage-lambda genomic library was made from a mixture of day 18 and day 19 fetal thymocyte DNAs and screened with the pG4.1 probe. Four distinct types of rearrangement were evident (Fig. 2).

1. *D-D-J* joining. As illustrated in Fig. 2 and the corresponding sequence in Fig. 3, this rearrangement is due to the joining of $D_{\delta 1}$, $D_{\delta 2}$ and $J_{\delta 1}$. Such rearrangements have been suggested to occur in the TCR β locus^{20,21}, but no direct evidence has ever been found.

2. *V-D* joining. Analysis of the λ M21 clone reveals a V region (V_{M21}) joined to $D_{\delta 2}$ but with no rearrangement to $J_{\delta 1}$ (Figs 2 and 3). This VD rearrangement without DJ joining has also not been described previously for any antigen-receptor gene. Although either *V-J* or *D-J* rearrangements seem allowable under the 12/23 rule, *D-J* joints are found frequently in immature B cells and seem to be the first step in the assembly of a complete immunoglobulin heavy-chain gene^{22,23}. Similarly, analysis of TCR β -chain rearrangement in developing thymocytes suggests that $D_{\beta} \rightarrow J_{\beta}$ joining precedes other types of rearrangements during ontogeny²⁴. In contrast, Southern analysis of developing thymocytes show that this particular rearrangement of V_{M21} to $D_{\delta 2}$ corresponds to one of the major rearranged bands for TCR δ in both fetal thymocytes (data not shown) and hybridomas (see next section). This suggests that *D-J* joining is not necessarily the first event in the formation of all antigen-receptor genes containing D segments.

3. *VDJ* joinings. Five of the phage clones isolated have $VD_{\delta 2}J_{\delta 1}$ rearrangements. Restriction-enzyme mapping and DNA-sequence analysis show that three of these (clones M2, M9 and M16) have rearrangements that involve the same V-region gene segment (V_{M21}) described above. This sequence, as well as a second V region (V_{M11}), show little similarity with any known V_{α} -gene segments¹³. In contrast, the third V segment (V_{M23}), shows more than 90% nucleotide-sequence similarity with the previously described V_{α} gene, TA27 (ref. 13). Overall

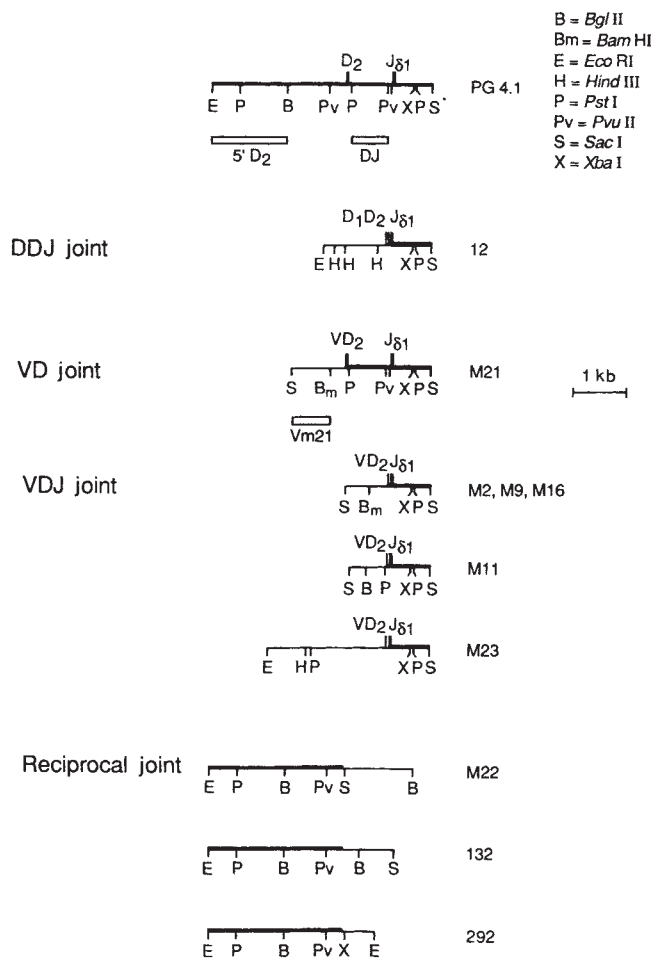


Fig. 2 Restriction maps of pG4.1 and rearranged clones isolated from a 18- and 19-day fetal thymocyte genomic-DNA library.

Methods. A bacteriophage-lambda genomic library⁴⁷ was constructed from a mixture of day-18 and day-19 fetal thymocyte DNA and screened with the pG4.1 probe. From 1.4×10^6 original clones (representing the equivalent of eight mouse genomes), 14 clones were isolated. Three of them are in the germline configuration. The rearranged region of the remaining ten phage clones were subcloned and analysed further. The nature of the rearrangements is determined by restriction-enzyme mapping and DNA-sequence analysis. Relevant sequences are presented in Fig. 3. The name of each clone appears after its restriction map. The filled areas indicate sequence similarity between the clones and pG4.1 as determined by restriction-enzyme mapping and DNA-sequence analysis. M2, M9 and M16 have the same restriction maps with respect to the enzymes used here, therefore only one map is presented. But the sequences at the junctions are all different and are reported separately in Fig. 3. The probes used for the Southern analysis of rearrangement (Fig. 4) are: 5' $D_{\delta 2}$, a 1.5 kb *Eco*RI to *Bgl*III fragment; *DJ*, a 0.8 kb *Pst*I to *Pvu*II fragment and V_{M21} , a 0.7 kb *Sac*I to *Bam*HI fragment. Their positions are indicated by open boxes.

this suggests that there may be differences between the V_{δ} and V_{α} repertoires, although a systematic study of V_{α} usage in fetal thymocytes has not been reported. We also observe a wide variation in the number of N-linked glycosylation sites in the V regions described here (0-3). This may have a significant effect on the apparent relative molecular mass of TCR δ chains employing different V segments.

In these fetal thymocyte rearrangements, all of the sequences between V and J can be accounted for by $D_{\delta 2}$ alone, with no discernible N-region contribution²⁵. The occurrence of N-region addition can be correlated with the enzyme terminal-deoxynu-

DDJ	GGTGATGCCAAATCCAAGGGAAGAACAAAGGG <u>TGTTTTGT</u> ACGGCTGTGTTT <u>CACCTGTG</u> GTGGCATAT ATCGGAGGGATAC ACAAACCTCGTCTTTGGACAAGGAACCCCAAGTGACTGTGGAACCAA	12nt	D _{δ1}	D _{δ2}	J _{δ1}	12
VD	Asp Val Tyr Leu Gln Pro Val Ala Lys Thr Phe Thr Val Val Ala Gly Asp Pro Ala Ser Phe Tyr Cys Thr Val Thr Gly Gly Asp Met Lys Asn Tyr His GAT GTA TAT TTG GAA CCA GTT GCC AAA ACT TTT ACT GTT GTA GCT GGG GAT CCT GCC TCC TTC TAC TGC ACT GTA ACA GGA GGG GAC ATG AAG AAT TAT CAT	+1	←V→	20		M21
VDJ						M2
						M9
						M16
	Gly Val Asn Thr Gln Met Leu His Gln Ser Pro Gln Ser Leu Thr Ile Gln Glu Gly Asp Glu Val Thr Met Ser Cys GGG GTC AAC ACC CAA ATG TTG CAT CAG AGT CCT CAG TCT CTG ACA ATC CAA GAA GGA GAT GAA GTC ACC ATG AGC TGC	+1	←V→	20	CHO Asn Leu Ser Thr Ser Leu Tyr Ala AAT TTA TCG ACA TCT TTA TAT GCC	M11
	Ala Gly Ser Asn Val Ala Gln Lys Val Ile Gln Val Trp Ser Thr Thr Ser Arg Gln Glu Gly Glu Lys Leu Thr Leu Asp Cys Ser Tyr Lys Thr Ser Gln Val Leu Tyr His Leu GCA GGA TCT AAT GTG GCC CAG AAA GTG ATT GAG GTC TGG TCA ACA ACA AGC AGG CAG GAG GGA GAA AAA CTC ACA CTG GAC TGT TCA TAT AAG ACA AGT CAG GTC TTA TAC CAT CTT	+1	←V→	20		M23
	Met Ser Trp Tyr Lys Lys Asn Gly Thr Asn Ala Leu Phe Leu Val Tyr Lys Leu Asn Ser Asn Ser Thr Asp Gly Gly Lys Ser Asn Leu Lys Gly Lys Ile Asn Ile Ser Lys Asn Gln ATG AGC TGG TAT AAG AAG AAT GGA ACT AAT GCT CTG TTT TTA GTA TAC AAG CTA AAT AGC AAT TCT ACT GAT GGT GGA AAG AGC AAC CTC AAA GGG AAA ATT AAC ATT TCA AAA AAT CAG	40	←V→	60	CHO	M21
						M2
						M9
						M16
	Leu Leu Trp Tyr Arg Gln Gly Asp Asp Gly Ser Leu Val Ser Leu Val Thr Leu Gln Lys Gly Gly Asp Glu Lys Ser Lys Asp Lys Ile Thr Ala Lys Leu Asp Lys Lys Met Gln Gln TTA CTC TGG TAT AGG CAG GGG GAT GAT GGC AGC CTC GTC TCC TTG GTG ACA TTA CAG AAA GGA GGA GAT GAG AAA AGT AAG GAC AAA ATA ACC GCT AAT CTG GAT AAG AAA ATG CAG CAA	40	←V→	60		M11
	Phe Trp Tyr Lys His Leu Leu Ser Gly Glu Met Val Leu Leu Ile Arg Gln Met Pro Ser Thr Ile Ala Ile Glu Arg Ser Gly Arg Tyr Ser Val Phe Gln Lys Ser Arg Lys Ser TTC TGG TAC AAG CAC CTT CTT AGT GGA GAG ATG GTT TTG CTT ATT CGA CAA ATG CCT TCT ACT ATT GCA ATA GAG AGG AGC GGC CGC TAT TCT GTA GTC TTC CAG AAA TCA CGC AAA TCC	40	←V→	60		M23
	Phe Ile Leu Asp Ile Gln Lys Ala Thr Met Lys Asp Ala Gly Thr Tyr Tyr Cys Gly Ser Asp TTT ATA CTC GAC ATT CAG AAG GCA ACA ATG AAA GAC GCT GGT ACG TAC TAC TGT GGG TCA GAT AT	80	←V→	95	D _{δ2}	M21
				22nt	CGGAGGGATACGAG <u>CACAGTG</u> TTGCAACCCCATAGGGACCTGT <u>ACAAAACT</u> GCAGGGTGAAGAA	
					Asp Lys Leu Val Phe Gly Gln Gly Thr Gln Val Thr CC GAC AAA CTC GTC TTT GGA CAA GGA ACC CAA GTG ACT	M2
						M9
						M16
	Ser Ser Leu Gln Ile Gln Ala Ser Gln Pro Ser His Ser Gly Thr Tyr Leu Cys Gly Gly Lys AGT TCC CTG CAG ATC CAA GCC TCC CAG CCC AGC CAC TCA GGC ACT TAC CTC TGT GGA GGG AAA	80	←V→	95	Thr	M11
	Ile Ser Leu Val Ile Ser Thr Leu Gln Pro Asp Asp Ser Gly Lys Tyr Phe Cys Ala Leu Trp Glu ATC AGC CTT GTC ATT TCA ACC TTA CAA CCA GAC GAT TCG GGA AAG TAT TTC TGT GCT CTC TGG GAG CT GGAGGGATACG	80	←V→	95		M23
Reciprocal Joints	GACACGTGATACAAAGCCAGGGAA <u>GGTTTTTGC</u> AAAGCTCTGTAG <u>CACCGTG</u> <u>CACCTGTG</u> ACAGACAGGGCTGCACCTGT <u>ACAAAATCT</u> GCCCGATGGGGAGGGGAAGCAGAGCAT	12nt		20nt		M22
	GACACGTGATACAAAGCCAGGGAA <u>GGTTTTTGC</u> AAAGCTCTGTAG <u>CACCGTG</u> <u>CACAGTG</u> GTGCACAAGCACCTGCACCTGTGTC <u>ATAAACCCA</u> CAGCTGCTGTGTGCCCTACCTGTC	12nt		25nt		132
	GACACGTGATACAAAGCCAGGGAA <u>GGTTTTTGC</u> AAAGCTCTGTAG <u>CACCGTG</u> <u>CACCGTG</u> ACACAGGCAACAGAGGGGTCTA <u>ACAGAAAC</u> TCAGATCTCAGCATCTGTGCASATG	12nt		23nt		292

Fig. 3 Sequence analysis of rearranged TCR δ genes in 18–19-day fetal thymocytes. Sequences of rearranged δ genes were analysed as described in the legend to Fig. 1. The restriction maps of each clone are shown in Fig. 2. V, D and J contributed sequences are indicated as well as possible N-linked glycosylation sites (denoted 'CHO'). The 3' end of the V region was chosen somewhat arbitrarily by lining it up with published germline V_{α} genes. The analysis of the junctional sequence is further complicated by the three nucleotides, ATC, shared by $D_{\delta 1}$ and $D_{\delta 2}$ and the three nucleotides, TAC, shared by $D_{\delta 2}$ and the 5' end of $J_{\delta 1}$. Therefore, it is not possible to determine precisely where the VD and DJ junction are, nor if $D_{\delta 1}$ is involved. Both M2 and M16 were only sequenced from the *Bam*HI site (Fig. 2). A gap was introduced in the M11 sequence to maximize the similarity.

cleotidyl transferase (TdT)²⁶. This enzyme is noticeably less abundant in fetal (versus adult) thymocytes in both the avian and murine systems^{27,28}. Four out of five *V-D-J* rearrangements analysed here result in out-of-frame joints. This is in contrast to our results with cDNA clones isolated from adult double-negative thymocytes¹⁹, and many of these non-functional genes may come from cells already in the $\alpha:\beta$ -bearing phase.

4. Joining of reciprocal recognition heptamers. The three rearrangements found in λ M22, λ 192 and λ 232 represent the 'reciprocal joint' type of recombination first described in immunoglobulin genes²⁹. They consist of the two joining signals fused to each other at their (formerly) coding-region proximal borders. In all three cases, the putative joining-signal 5' of the *D*₂ element (GGTTTTTGG...[13]...CACCGTG) is fused with signal sequences having a 23-base-pair spacing that can be found at the 3' end of the *V* or *D* elements, forming a head-to-head structure of the two heptamers. All the joinings are precise and do not involve the addition or deletion of nucleotides. Three models have been proposed to explain this phenomenon; reintegration of deleted sequences, unequal sister-chromatid exchange and inversion¹³. More recently, products of reciprocal joints from both the TCR α and β locus have been found in circular DNA isolated from thymocytes^{30,31}. An example of a *V_β*-gene segment rearranged and expressed by inversion has also been documented³². Furthermore, analysis of β -gene rearrangement in some of the T cells show a partial duplication of the *J_β-C_β* locus. This latter finding seems consistent only with the unequal sister-chromatid exchange model¹³. The high frequency with which we observe this reciprocal-joint type of rearrangement raises the interesting question as to what are the mechanisms of rearrangement in *V*-gene assembly at the δ locus.

Survey of fetal thymocyte hybridomas

To analyse TCR δ rearrangements at a clonal level throughout T-cell ontogeny, and to be able to compare them with *V*-region assembly in the β -chain and γ -chain loci, we analysed a series of hybridomas made by fusing mouse fetal thymocytes from day 14, 15, 16 and 18 of gestation with BW5147 (ref. 24), which has deleted the δ gene from both chromosomes. Southern blots of hybridoma DNAs digested with *Eco*RI were screened with four probes designed to detect each class of rearrangement at the *J_δ* locus. These probes are indicated in Figs 1 and 2 and described in Fig. 4. *Hind*III digested DNAs were probed with a 2 kb *Sac*I-*Sac*I fragment containing the 3' half of *J_δ* and a 2.8 kb *Eco*RI fragment located just 5' of *J_δ* to detect rearrangement at this locus. When rearrangements did not fall into any of the categories discussed above, Southern blots were prepared using restriction enzymes other than *Eco*RI or *Hind*III, and probes generated from the cosmid clones were used to determine the nature of the rearrangements.

Six of the 23 day-14 hybrids show rearrangements of *J_δ* (Fig. 4). No rearrangements of *J_δ* were observed. This is consistent with our previous finding that *J_δ* rearrangements can be detected in day-14 fetal-thymocyte DNA using the pG4.1 probe. As summarized in Table 1, only one of the six rearrangements is DJ joining in nature; four involve *V_{M21}* and four involve rearrangements 5' of the *D_δ* element. This correlates with the *V*→*D* rearrangement discussed previously and contrasts with the situation in immunoglobulin heavy chain and TCR β -chain gene rearrangement where DJ joinings are the first to appear during ontogeny and *V*→*D* joinings have never been reported^{22,24}. The apparent preferential usage of *V_{M21}* in these rearrangements does have a parallel in immunoglobulin gene assembly in that pre-B-cell lines greatly favour rearrangements with two of the most *J_H*-proximal *V_H*-gene segments^{33,34}.

Studies of β - and γ -gene rearrangement in this set of hybridomas show that only four have TCR- γ or TCR- β rearrangements, two of which (number 3 and number 13) have only *DJ_β* rearrangements, one (number 1) has only a *C_γ* rearrangement, and one (number 12) shows both *C_γ* and *DJ_β* rearrange-

ments; all of the other hybridomas show germline configurations for both β and γ loci^{24,35}. The fact that six of the day-14 hybrids have TCR δ rearrangements suggests that rearrangement at this locus occurs at the same time, if not slightly earlier, than in either TCR β and TCR γ .

Similar analyses were carried out on 22 day-15 fetal hybridomas (as summarized in Table 1). Half of the hybrids are in the germline configuration, with the other half exhibiting more complex rearrangements than in the day-14 fetal hybridomas. In addition to those observed previously, namely the *D_δ**J_δ**1*, *D_δ**D_δ**2*, *V_{M21}**D_δ* and *V_{M21}**D_δ**J_δ**1* types, rearrangements involving at least four, possibly six, distinct sequences were observed, all of which could be *V* regions other than *V_{M21}*. One of these joins to *D_δ* without a *D_δ**J_δ**1* deletion, suggesting that *V* regions other than *V_{M21}* can be involved in *V*→*D* joinings. Two hybridomas show rearrangements at the *J_δ* locus resulting in the deletion of *J_δ**1*. One hybrid has a *V_{M21}**DJ_δ**1* rearrangement on one chromosome, and on the other a *V_{M21}*-gene segment has rearranged into a region 3' of *C_δ* (~65 kb 5' of *C_α*) where functional *J_α*s have been identified³⁶. This suggests that *V*→*J_α* rearrangements could occur even at this very early stage of development.

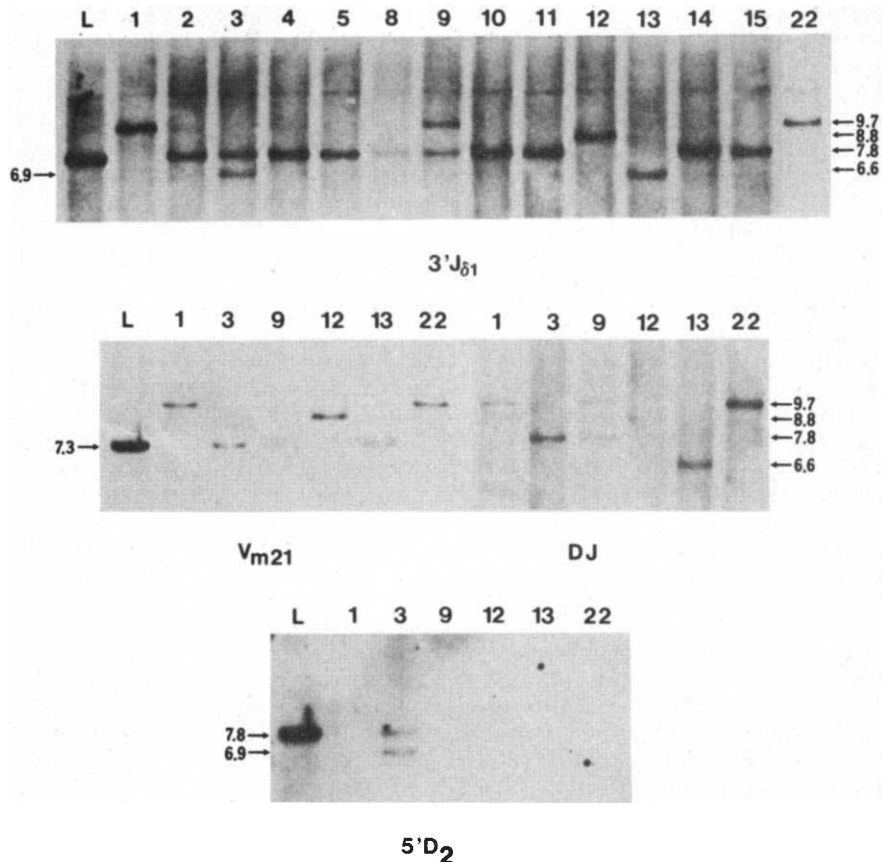
Twenty-four day-16 and eighteen day-18 hybridomas were analysed similarly (see Table 1). The day-16 hybrids are less complex than the day-15 hybrids, the instance of *V*→*D* rearrangement is higher (7/24 versus 3/20) and five of them involve *V_{M21}*. Again there are no rearrangements at the *J_δ* locus within this set of hybrids. By day 18, most of the hybrids have δ -gene rearrangements on both chromosomes and these appear to be *VDJ_δ* in nature. Several *V*→*J_α* rearrangements could also be identified in the day-18 hybrids, consistent with the appearance of significant numbers of $\alpha:\beta$ -bearing cells at this time, and these were excluded from the table. It is significant that with the 5' *D_δ*-region probe we could find no evidence of reciprocal joins in these hybridomas, despite the three clones isolated from the 18- and 19-day genomic library discussed previously and the prominent hybridization of fetal thymocyte DNAs on Southern blots probed with the 5'-*D*-region probe (data not shown). One possibility is that such products are residing on circular DNA molecules which are the apparent deletion products from both the α - and the β -locus thymocyte DNA (refs 30, 31). Such recombination intermediates could be lost from hybridomas which have been grown in culture, but are sufficiently numerous to be observed in thymocytes. Alternatively, it is possible that only a subset of the *V* regions are joined by inversion to the *DJ* element during *V*-region assembly, and the *V* regions we analysed here are a different subset, such as has been postulated in the *Igκ* locus³⁷. Consistent with this hypothesis, Southern analysis of a panel of adult dull-CD5 double-negative hybridomas showed that several of them have bands that could correspond to such products. A third formal possibility is that the products of these reciprocal joints are lost from the chromosomes of these hybridomas by an as yet unidentified secondary rearrangement.

Analysis of the *V_{M21}* hybridization pattern shows that it is in the germline configuration in cells with *D_δ**J_δ**1* or *D_δ**D_δ**J_δ**1* rearrangements, but is deleted from all other chromosomes where rearrangements other than *DJ* are detected. This suggests it is located 5' to the *D* elements and that it is the most *DJC*-proximal *V* region studied here.

TCR δ and TCR γ

As TCR δ pairs with TCR γ , it is relevant to determine how their rearrangements correlate with each other. In the analysis of the δ locus, we score each individual hybrid according to its most advanced rearrangement. In a given hybrid, if one chromosome shows a *VDJ* rearrangement, then it is categorized as a *VDJ* hybrid, regardless of the status of the other chromosome. Rearrangements of the γ locus in these hybridomas were described previously³⁵.

Fig. 4 Southern analysis of δ rearrangements in day-14 fetal thymocyte hybridomas. The first panel shows 14 of the 22 day-14 hybridomas analysed with the 3' $J_{\delta 1}$ probe, a 2 kb *SacI* fragment located 3' of $J_{\delta 1}$, used to detect rearrangement in general, and compared with liver DNA (L). Hybridomas NDTE-1, 3, 9, 12, 13 and 22 (denoted by their numbers alone) showed evidence of rearrangement and were analysed further. That these rearrangements correspond to $V_{M21} \rightarrow D_{\delta 2}$ events in hybrids 1, 9 and 22, and a $V_{M21}DJ_{\delta 1}$ joining in hybrid 12 was determined by probing the blots with the V_{M21} probe, a 0.8 kb *Bam*HI to *SacI* fragment which contains mostly the 5' flanking-region of V_{M21} and detects rearrangements involving this V region. Analysis of λ clones show that it detects a 9.7 kb band in a $VD_{\delta 2}$ rearrangement and a 8.8 kb band in a $VDJ_{\delta 1}$ joining situation. It should be noted that Southern analysis will not show whether or not $D_{\delta 1}$ is involved. The DJ probe is a 0.7 kb *Pst*I-*Pvu*II fragment located between $D_{\delta 2}$ and $J_{\delta 1}$ used to detect $D_{\delta 2}J_{\delta 1}$ joining (in which case the region is deleted). The DJ probe also detects the 6.6-kb band in hybrid 13; the same band hybridized with a 1.9 kb *Eco*RI fragment containing $D_{\delta 1}$ (Fig. 1) (data not shown) suggesting that this rearrangement is a $D_{\delta 1}D_{\delta 2}$ joining with $J_{\delta 1}$ in the germline position. The rearrangement in hybrid 3 is designated as a $D_{\delta 2}J_{\delta 1}$ joining because it is detected by the 5' D_2 probe, but no hybridization is observed with the $D-J$ probe. The 5' D_2 probe is a 1.25 kb *Eco*RI-*Bgl*II fragment which detects rearrangements involving $D_{\delta 2}$, such as $D_{\delta 2}J_{\delta 1}$ rearrangement and reciprocal joining of the heptamer sequences. (The 3' $J_{\delta 1}$ is described in Fig. 1 and the other probes are described in Fig. 3). Sizes of the different bands are given in kilobases. **Methods.** Aliquots of 8 μ g *Eco*RI-digested genomic DNA from liver (L) or day-14 fetal thymocyte hybridomas were electrophoresed into a 0.8% agarose gel and blotted on to nitrocellulose. All the probes were radiolabelled by hexamer priming⁴⁸. The blots were hybridized in 1 M NaCl, 50 mM sodium phosphate, pH 6.5, 50% formamide, 100 μ g salmon-sperm DNA per ml, 4 $\times$ Denhardt's solution and 10% dextran sulphate at 40 $^{\circ}$ C and washed in 2 $\times$ SSPE and 0.2 $\times$ SSPE at 55 $^{\circ}$ C. SSPE in 3.6 M NaCl, 0.2 M sodium phosphate, pH 7.4, 0.02 M EDTA.



A comparison of the γ - and δ -loci rearrangements is shown in Table 2. It appears that the majority (seven) of the day-15 hybrids have rearrangements only at the $C_{\gamma 2}$ locus and that these are for the most part accompanied by potentially functional TCR- δ rearrangements (VDJ). In contrast, among the day-16 hybridomas which show rearrangement at γ loci, most of them show rearrangements at the $C_{\gamma 1}$ locus only. Examination of the status of δ locus in the $C_{\gamma 1}$ -only category shows that they either have the germline form (three) or non-functional rearrangements (six). This is consistent with the serological studies indicating that γ -chain proteins in developing fetal thymocytes are primarily derived from the $C_{\gamma 2}$ locus¹¹.

Discussion

Our analysis of TCR δ -locus rearrangements in fetal thymocytes shows that they can be quite diverse. A number of V -region segments can be used, as well as at least two D regions which can be used together or separately and can be read in all three reading frames. One striking difference in the assembly of the δ locus in early thymocytes as compared with that of immunoglobulin heavy chain is the fact that the initial rearrangements seem to centre around the $D_{\delta 2}$ element. Deletion of the region 5' of $D_{\delta 2}$ (the $VD_{\delta 2}$ and $D_{\delta 1}D_{\delta 2}$ types of rearrangement) is at least as frequent as $D_{\delta 2} \rightarrow J_{\delta 1}$ joinings, despite the close proximity of $D_{\delta 2}$ and $J_{\delta 1}$. ($J_{\delta 1}$ is 0.9 kb 3' to $D_{\delta 2}$, $D_{\delta 1}$ is 9 kb 5' to $D_{\delta 2}$, and V regions are located no closer than 30 kb 5' to $D_{\delta 2}$). It is thus difficult to reconcile our data with the 'one-dimensional scanning model' suggested for a putative recombinase during the early stages of DJ and V to DJ joining in

immunoglobulin heavy-chain V -gene assembly³⁸. In fact, recent experiments have now convincingly eliminated tracking models for recombination in prokaryotic systems altogether. Instead, it has been suggested that it is the local geometry of the two sites that is important, not the presence or absence of a continuous DNA path connecting them^{39,40}. Similarly, in studies of gene regulation it appears that proteins bound to 'regulatory elements' can interact with distant target sites by simple collision. Thus, the control of both recombination and transcription seems to depend more on protein-protein interactions than on the linear arrangement of binding sites on the DNA *per se*⁴¹.

The interesting skewing of TCR-gene rearrangements at different days of fetal thymocyte development as indicated by the hybridoma analysis (Table 2) points out the dynamic nature of the fetal thymocytes populations. On day 15 most of the hybrids have $VDJ_{\delta 1}$ and $VJ_{\gamma}C_{\gamma 2}$ rearrangements. On day 16, however, most of the hybrids have non-functional δ rearrangements together with $VJ_{\gamma}C_{\gamma 1}$ joinings, a γ -chain gene whose product is not seen in early thymocytes¹¹. As the γ and δ rearrangements are not cumulative, the day-15 hybridomas with TCR rearrangements cannot be the precursors of day-16 hybridomas. Rather, a major change in the component subpopulations seems to be taking place. This is not completely surprising as the total number of cells in the thymus increases exponentially during this period of development⁴². The selective outgrowth and rearrangement of some subset of the day-15 cells, migration of lymphocyte precursors from the fetal liver with different proliferative potentials⁴³, as well as selective death or export of the day-15 rearranged population, are all potential explanations

Table 1 Rearrangement of the TCR δ locus in fetal thymocyte hybridomas

Time of gestation	Number of hybridomas analysed	$D_{\delta 1}D_{\delta 2}$	$VD_{\delta 2}$	Types of rearrangement			Other	
				$V_{M21}D_{\delta 2}$	$D_{\delta 2}J_{\delta 1}$	$V_{M21}DJ_{\delta 1}$		
14	21	1	0	3	1	1	0	(6)
15	20	1	1	2	6	3	5*	(18)
16	24	0	2	5	5	4	1	(17)
18	17	0	0	0	6	2	19†	(27)

Each chromosome that shows rearrangement detected by the probes described in the text was categorized. Numbers in parenthesis at the end of the row indicate the total number of chromosomes that show rearrangement in all the hybridomas at a given stage. Only one type of rearrangement is observed for any given chromosome. Chromosomes that have lost C_{δ} (Fig. 1) hybridizable bands are excluded. This could be due to either the loss of the chromosome or to rearrangement in the α locus.

* Two of the chromosomes in this panel of day-15 hybridomas show rearrangement at the $J_{\delta 2}$ locus.

† All 19 rearrangements are at the $J_{\delta 1}$ locus. Rearrangements at the $J_{\delta 2}$ locus were not determined.

Table 2 Correlation of TCR γ and δ rearrangements in fetal ontogeny

Days of gestation	14			15			16		
	Germline only	VD_{δ} or DJ_{δ}	VDJ_{δ}	Germline only	VD_{δ} or DJ_{δ}	VDJ_{δ}	Germline only	VD_{δ} or DJ_{δ}	VDJ_{δ}
Germline only	14	4	0	11	1	1	4	2	1
$C_{\gamma 1}(C_{\gamma 10.8})$ only	0	1	0	0	0	2	3	6	0
$C_{\gamma 2}(C_{\gamma 13.6})$ only	0	0	1	0	2	5	0	0	2
$C_{\gamma} + C_{\gamma 2}$	0	0	0	0	1	0	0	2	1

Each hybridoma was scored only once according to their rearrangement at the δ locus as described in the text. The rearrangements at the γ locus have been described³⁵. The probe used does not cross-hybridize with $C_{\gamma 4}$ (ref. 49). The nomenclature for the γ genes is $C_{\gamma 1} = C_{\gamma 10.5}$; $C_{\gamma 2} = C_{\gamma 13.4}$; $C_{\gamma 3} = C_{\gamma 7.5}$, where 10.5, 13.4 and 7.4 correspond to the 10.5, 13.4, and 7.5 kb *Eco*RI fragments in the BALB/c genome. It should be noted that none of the hybrids shows a deletion of C_{γ} -gene containing germline bands without corresponding rearrangements at that same locus.³⁵.

for this change in the composition of the thymus. That embryonic $\gamma:\delta$ cells can migrate out of the thymus is suggested by the finding that at least two of the $\gamma:\delta$ -bearing dendritic epithelial T-cell lines use $V_{\gamma 3}$ for the γ chain⁴⁴ and V_{M21} as the V region for the δ chains (Y.C., unpublished data and J. Allison, personal communication). Both V regions are found rearranged and expressed in abundance in fetal but not in adult cells.

These data also suggest that the relative order of TCR- γ rearrangements are different between different T-cell populations. Most of the day-15 hybrids have rearrangements of the $C_{\gamma 2}$ locus coupled with potentially functional δ genes, whereas day 16-hybrids show mostly $C_{\gamma 1}$ rearrangements with non-functional (VD or DJ) δ rearrangements. In contrast, many day-17 and adult dull-CD5, double-negative cell hybridomas have rearrangements both at the $C_{\gamma 1}$ and $C_{\gamma 2}$ locus, suggesting that there is coordinate control of the γ and δ rearrangement and expression during different stages of T-cell development.

We wish to acknowledge the NIH and the Howard Hughes Medical Institute for financial support and Kathy Redman for preparation of the manuscript. M.I. was supported, in part, by a fellowship from the Fulbright Foundation, D.A.W. is supported by the Medical Scientist Training Program (GM07365), J.F.E. is a Centennial Fellow of the Medical Research Council of Canada and M.M.D. is a Scholar of the Pew Foundation.

Note added in proof: We have determined the germline sequence of V_{M21} and found the assignment of the 3' end as postulated in Fig. 3 to be correct. The $D_{\delta 1}$ and $D_{\delta 2}$ sequences from B10.A mice were determined by the polymerase chain reaction (R. K. Saiki *et al.* *Nature* **324**, 163–166, 1987) and direct sequencing. The $D_{\delta 1}$ sequence of B10.A is ATGGCATATCA. The $D_{\delta 2}$ sequence of B10.A is identical to that in BALB/c. Thus, the nature of the gap in the $V_{\delta 12}$ $D_{\delta 2}$ sequence (Fig. 1c) is not clear and this issue is under investigation.

Received 10 September; accepted 4 November 1987.

- Chien, Y., Iwashima, M., Kaplan, K. B., Elliott, J. F. & Davis, M. M. *Nature* **328**, 677–682 (1987).
- Lindsten, T., Fowlkes, B. J., Samelson, L. E., Davis, M. M. & Chien, Y. *J. exp. Med.* **166**, 761–775 (1987).
- Born, W. *et al.* *Nature* **330**, 573 (1987).
- Loh, E. *et al.* *Nature* **330**, 569 (1987).
- Bonyhadi, M., Weiss, A., Tucker, P. W., Tigelaar, R. E. & Allison, J. P. *Nature* (in the press).
- Brenner, M. B. *et al.* *Nature* **322**, 145–149 (1986).
- Bank, I. *et al.* *Nature* **322**, 179–181 (1986).
- Lew, A. M. *et al.* *Science* **234**, 1401–1405 (1986).
- Borst, J. *et al.* *Nature* **325**, 683–688 (1987).
- Moingeon, P. *et al.* *Nature* **325**, 723–726 (1987).
- Pardoll, D. M. *et al.* *Nature* **326**, 79–81 (1987).
- Bluestone, J. A., Pardoll, D. M., Sharrow, S. O. & Fowlkes, B. J. *Nature* **326**, 82–84 (1987).
- Kronenberg, M., Siu, G., Hood, L. E. & Shastri, N. A. *Rev. Immun.* **4**, 529–591 (1986).
- Dembic, Z., Bannwarth, W., Taylor, B. A. & Steinmetz, M. *Nature* **314**, 271–273 (1985).
- Saito, T., Weiss, A., Miller, J., Norcross, M. A. & Germain, R. N. *Nature* **325**, 125–130 (1987).
- Tonegawa, S. *Nature* **302**, 575–581 (1983).
- Early, P., Huang, H., Davis, M., Calame, K. & Hood, L. *Cell* **19**, 981–992 (1980).
- Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. *Nature* **286**, 676–683 (1980).
- Elliott, J. F., Rock, E. P., Davis, M. M. & Chien, Y. *Nature* (in the press).
- Kavaler, J., Davis, M. M. & Chien, Y. *Nature* **310**, 421–423 (1984).
- Siu, G. *et al.* *Nature* **311**, 344–350 (1984).
- Yaoita, Y. *et al.* *Nucleic Acids Res.* **11**, 7303–7316 (1983).
- Alt, F. W. *et al.* *EMBO J.* **3**, 1209–1219 (1984).

- Born, W., Yague, J., Palmer, E., Kappler, J. & Marrack, P. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2925–2929 (1985).
- Alt, F. W. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4118–4122 (1982).
- Landau, N. R., Schatz, P. G., Rosa, M. & Baltimore, D. *Molec. cell. Biol.* **7**, 3237–3243 (1987).
- Penit, C., Jotereau, F. & Gelibert, M. *J. Immun.* **134**, 2149–2154 (1985).
- Rothenberg, E. & Triglia, D. *J. Immun.* **130**, 1627–1633 (1983).
- Steinmetz, M., Altenburger, W. & Zachau, H. *Nucleic Acids Res.* **8**, 1709–1720 (1980).
- Fujimoto, S. & Yamagishi, H. *Nature* **327**, 242–243 (1987).
- Okazaki, K., Davis, D. D. & Sakano, H. *Cell* **49**, 477–485 (1987).
- Malissen, M. *et al.* *Nature* **319**, 28–33 (1986).
- Yancopoulos, G. D. *et al.* *Nature* **311**, 727–733 (1984).
- Perlmutter, R. M., Kearney, J. F., Chang, S. P. & Hood, L. *Science* **227**, 1597–1601 (1985).
- Born, W., Rathbun, G., Tucker, P., Marrack, P. & Kappler, J. *Science* **234**, 479–482 (1986).
- Winoto, A., Mjolsness, S. & Hood, L. *Nature* **316**, 832–836 (1985).
- Lewis, S., Rosenberg, N., Alt, F. & Baltimore, D. *Cell* **30**, 807–816 (1982).
- Yancopoulos, G. D. & Alt, F. W. *A. Rev. Immun.* **4**, 339–368 (1986).
- Boocock, M. R., Brown, J. L. & Sherratt, D. J. *Biochem. Soc. Trans.* **14**, 214–216 (1985).
- Craigie, R. & Mizuuchi, K. *Cell* **45**, 795–800 (1983).
- Gellert, M. & Nash, H. *Nature* **325**, 401–404 (1987).
- Ceredig, R., Dyalynas, D. P., Fitch, F. W. & MacDonald, H. R. *J. exp. Med.* **158**, 1654–1671 (1983).
- Jotereau, F., Heuze, F., Salomon-Vie, V. & Gascan, H. *J. Immun.* **138**, 1026–1030 (1987).
- Garman, R. D., Doherty, P. J. & Raulet, D. H. *Cell* **45**, 733–742 (1986).
- Chen, E. Y. & Seeburg, P. H. *DNA* **4**, 165–170 (1985).
- Henikoff, S. *Gene* **28**, 351–359 (1984).
- Chien, Y., Gascoigne, N. R. J., Kavaler, J., Lee, N. E. & Davis, M. M. *Nature* **309**, 322–326 (1984).
- Feinberg, A. & Vogelstein, B. *Analyt. Biochem.* **132**, 6–13 (1983).
- Iwamoto, A. *et al.* *J. exp. Med.* **163**, 1203–1212 (1986).

Evidence for interstellar SiC in the Murray carbonaceous meteorite

Thomas Bernatowicz*†, Gail Fraundorf§, Tang Ming‡, Edward Anders‡, Brigitte Wopenka*‡, Ernst Zinner*† & Phil Fraundorf§

* McDonnell Center for the Space Sciences, † Physics Department and ‡ Department of Earth and Planetary Sciences, Washington University, St Louis, Missouri 63130, USA
§ Monsanto Research Center, St Louis, Missouri 63167, USA
‡ Enrico Fermi Institute and Department of Chemistry, University of Chicago, Chicago, Illinois 60637, USA

Silicon carbide has been observed in the interstellar medium and in circumstellar shells. Although it might be expected to occur in primitive meteorites that contain other kinds of pre-solar material, it has not previously been found. We have identified silicon carbide in two separates from the Murray carbonaceous chondrite that are enriched 20,000-fold in isotopically anomalous neon (Ne-E) and xenon (Xe-S)¹. An accompanying letter² reports on the isotopic composition of the two separates. The SiC is present in the form of crystalline grains, 0.1–1 µm in size. Cubic and {111}-twinned cubic are the most common ordered polytypes observed so far. The anomalous isotopic composition of its carbon, nitrogen and silicon indicates a pre-solar origin, probably in the atmospheres of red giants. An additional silicon- and oxygen-rich phase shows large isotopic anomalies in nitrogen and silicon, also associated with a pre-solar origin.

In a continuing search for the carriers of isotopically anomalous noble gas components¹, a bulk sample of the CM chondrite Murray was treated with HF-HCl, oxidized with Cr₂O₇²⁻ and boiling HClO₄ to remove organic and graphitic carbon, and separated by grain size into five fractions (using sedimentation in a centrifuge)¹. Sample CF comprises the grain sizes between 300 and 2,000 Å (28 p.p.m.). Another Murray separate, B2I, was obtained in a similar way, except that the size separation was achieved by filtration. The coarsest fraction (>2 µm, 1,860 p.p.m.), consisting mostly of spinel, was treated with H₃PO₄ to dissolve spinel, followed by HClO₄ and HNO₃ to dissolve graphitic carbon, yielding residue CJ (30 p.p.m.).

Samples CF and CJ have very high concentrations of Xe-S and Ne-E(H), respectively^{1,2}, and sample CJ shows substantial amounts of silicon by energy-dispersive X-ray (EDX) analysis¹, suggesting the presence of SiC. As preliminary ion probe measurements indicated more carbon than expected from SiC and lower ¹³C/¹²C ratios than had been found in CF (see ref. 2), CJ was oxidized for 30 min in an oxygen plasma discharge. The resulting residue, CJa, as well as residues CJ and CF, were studied by scanning electron microscope (SEM), ion microprobe, transmission electron microscope (TEM) and laser Raman microprobe analysis. Thin-specimen EDX and electron energy loss (EELS) analyses were made in the TEM, and SEM studies used an ultra-thin-window EDX detector, allowing the detection of elements with $Z \geq 6$.

In the SEM, CF appears as a fine-grained mixture, consisting mostly of spinel and Si-rich grains of 1,000–2,000 Å diameter in a ratio of ~10:1. Figure 1a shows an SEM micrograph of grains from Murray residue B2I, which is similar to CF but depleted in spinel. Sample CJa consists mostly of large (up to 20 µm) chromite crystals, accompanied by fine-grained Si-rich material (Fig. 1b) and occasional Mg-Al spinels. The EDX spectrum of the Si-rich material contains peaks for silicon and carbon, and also a prominent oxygen peak (but with an O/Si ratio less than that of SiO₂).

Silicon carbide was identified in residue CJa by laser Raman spectroscopy. Figure 2 shows the Raman spectrum obtained from a Si-rich area of a ~10-µm agglomerate. The prominent peak at ~795 cm⁻¹ is due to the transverse optical (TO) mode of SiC; the Raman shift of this mode is independent of polytype and propagation angle^{3,4}. In contrast, the shift of the longitudinal optical (LO) mode is known to be variable. There is an indication of the LO peak at ~930 cm⁻¹, much weaker than the TO peak. It has been shown previously⁵ that the intensity ratio of the two peaks depends greatly on the fabrication method and irradiation/heat treatment of SiC.

Further work showed that the separate CF also contains SiC crystals. TEM observations of CF reveal the existence of three major phases: (1) single crystals of Mg(Al,Cr)₂O₄ spinel with euhedral to rounded surfaces, ranging from 100 to several thousand angstroms in size; (2) a polycrystalline 'paste' of face-centred cubic (f.c.c.) diamond crystals ranging from 10 to many tens of angstroms in size; and (3) rather irregular twinned and/or faulted SiC crystals ranging from several hundred to

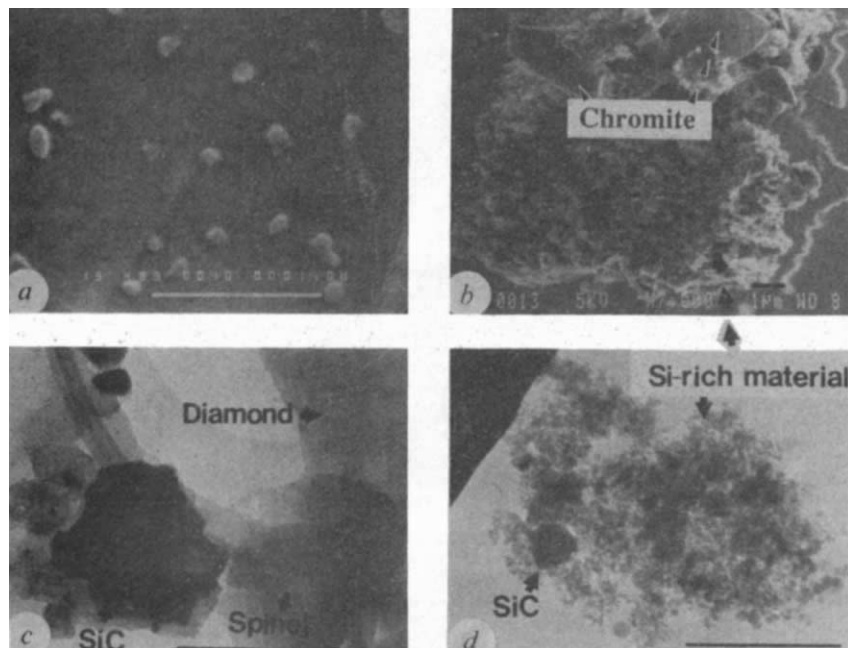


Fig. 1 a, SEM micrograph of Murray separate B2I, showing SiC and spinel grains. b, SEM micrograph of an agglomerate from Murray separate CJa, showing chromite crystals and attached fine-grained Si-rich material. c, TEM micrograph of separate CF, showing SiC, diamond and spinel grains. d, TEM micrograph of CJa, showing a SiC crystal and fine-grained Si,O-rich amorphous material. Scale bars: 1,000 Å for c; 1 µm for the other images.

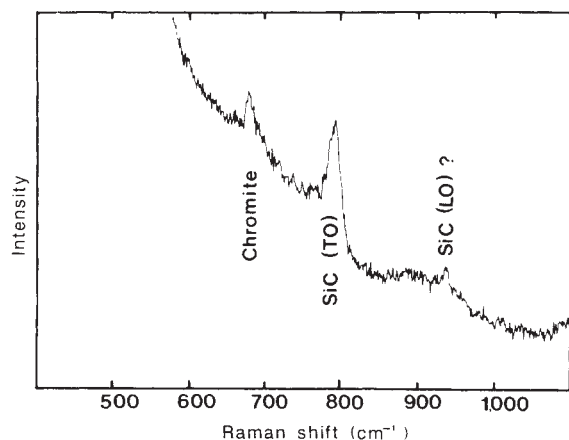


Fig. 2 Raman spectrum from a CJa agglomerate, showing a prominent TO (transverse optical) peak for SiC; also seen is a peak originating from chromite and an indication of the SiC LO (longitudinal optical) mode.

several thousand ångströms in size. Figure 1c shows a bright-field image of an agglomerate from residue CF.

Electron diffraction and high-resolution TEM imaging show that the spinel crystals have the expected crystal structure. EELS spectra show the fine-grained diamond paste to consist mostly of carbon, with diffraction spacings characteristic of f.c.c. diamond. Silicon is often present at low ($\leq 2\%$ by mass) levels as well. A lattice image of one of these diamond crystals shows the angle between 2.06-Å $\{111\}$ planes to be $\sim 70^\circ$, thus providing single-crystal data confirming the cubic diamond structure. This diamond phase appears to be identical to the 'C δ ' diamonds identified previously⁶.

Silicon and carbon were the only elements detected in EDX and EELS spectra of the SiC crystals. Electron diffraction patterns parallel and orthogonal to the tetrahedral layering axis for SiC reveal cubic SiC twinned along multiple $\{111\}$ planes, and SiC with nine-layer (22 Å) repeats (possibly the 27R polytype), in association with the twinned cubic polytype⁷. Patterns from some crystals show reciprocal lattice streaks in the layering direction, with little evidence for discrete spots, indicating severe stacking disorder. Dark-field imaging confirms the existence of twin domains in the cubic SiC, and high-resolution imaging (Fig. 3) permits direct observations of lattice and moiré fringes associated with the twins. High-resolution images further show evidence for the expected columns of Si-C atom pairs along $\langle 110 \rangle$, arranged in tetrahedral layers which are ordered according to the cubic (3H) stacking sequence.

TEM observations of CJ and CJa also show the presence of SiC, as well as abundant chromite and occasional spinel. SiC crystals are of the same average size as in CF, although several crystals up to $1\text{ }\mu\text{m}$ in diameter were observed. Electron diffraction patterns, as from CF, show evidence for cubic SiC with $\{111\}$ twinning, and also SiC with disorder in the layering direction. In order to characterize the full lattice, zone-axis patterns from two $\langle 110 \rangle$ zones separated by 90° , and a $\langle 100 \rangle$ zone pattern between these directions, were obtained from one single-crystal region of a SiC grain (see Fig. 4). The entire three-dimensional reciprocal lattice from this region is identical in spacings and interspot angles to that of cubic SiC (within measurement errors of $\sim 1\%$ for spacings), except for some evidence of excess streaking in one (only) of the four $\{111\}$ layer directions.

In addition to SiC, both CJ and CJa contain an abundant, poorly crystallized Si-rich phase present as aggregated $100\text{--}1,000\text{-Å}$ nodules (Fig. 1d), which has large inferred isotopic anomalies in silicon and nitrogen². This phase was not seen in CF and, if present there, must be a minor fraction. EELS spectra show Si and O but no C and N, indicating that the last two

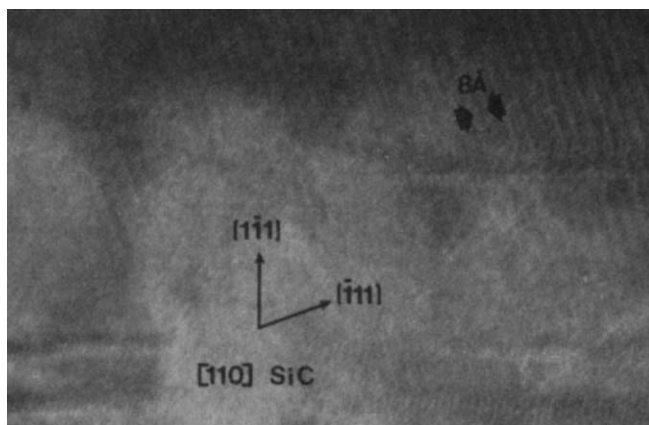


Fig. 3 $[110]$ zone-axis lattice image of a cubic SiC crystal with (111) twins, giving an edge-on view of (111) habit planes and the twin-parent interface. The 8-Å periodicity along $(\bar{1}11)$ in the parent crystal arises from moiré contrast with a superimposed (111) twin whose habit plane is nearly perpendicular to the $[110]$ electron beam direction.

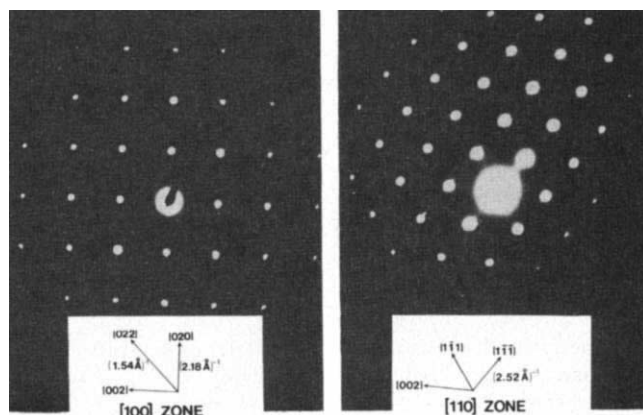


Fig. 4 Electron diffraction patterns along two zone axes separated by 45° , obtained from a CJ SiC crystal. The spacings, interplanar angles and interzone angles are characteristic of cubic SiC.

elements are less abundant by at least an order of magnitude. The same material with Si and O only is also found in CJ before plasma ashing, ruling out the possibility that it was formed from fine-grained SiC by the plasma oxidation process. In addition, plasma ashing of synthetic SiC produced no significant oxide layer. This material is apparently responsible for the substantial O signal seen in SEM EDX spectra of CJa agglomerates (Fig. 1b). Presumably it is a hydrolysis or oxidation product of some Si-bearing phase (nitride?) occluded by spinel, as silicon oxides, even if formed by some cosmic process, would have been destroyed by the HF treatment during sample preparation.

The Si/C elemental ratios measured in residue CF by ion microprobe analysis vary among different agglomerates, from almost zero (pure carbon phase) to values close to unity (SiC). There is no indication of any other Si-containing phase besides SiC. In contrast, Si/C ratios in CJa are >1 (ref. 2), confirming the presence of a second Si-bearing species. The ion probe was also used to look for the presence of nitrogen in SiC and in the amorphous Si-rich phase found in CJ. Nitrogen gives a strong CN^- signal in the presence of carbon². In CF the CN^-/C^- ratio indicates a nitrogen concentration of 1 wt % in SiC; in CJa the CN^-/C^- ratio varies among agglomerates, but is on average substantially higher than in CF. To check whether the additional nitrogen is associated with silicon, we measured the Si_2N^+ signal in different CJa agglomerates. The $\text{Si}_2\text{N}^+/\text{Si}^+$ signal ratio turned out to be comparable to that obtained from silicon with $\sim 2\text{ wt } \%$

of implanted nitrogen. Such a concentration would be consistent with the present N limit set by EELS for nitrogen in the Si, O-rich material seen in the TEM. The observation of large N and Si isotopic anomalies² in CJa, exceeding those observed in CF SiC, indicates that both elements are in the same phase in CJa—either in the amorphous Si₂O-phase (if derived from a Si-N mineral, it may have retained some nitrogen), or in a separate phase that we have not yet observed in the TEM.

The isotopic measurements reported in ref 2 reveal large C, N and Si isotopic anomalies in the SiC and the Si₂O-rich phases, indicating a pre-solar origin. Previous carbon isotopic measurements in CM chondrites have already indicated the existence of a refractory phase with high ¹³C/¹²C ratios⁸⁻¹⁰. Zinner and Epstein¹¹ tentatively proposed SiC as the carrier of heavy carbon in micrometre-size grains associated with oxide grains in Murchison. Raman microprobe analysis of a Leoville acid residue showed the presence of SiC but, because the silicon was isotopically normal, terrestrial contamination could not be ruled out¹². This work is thus the first unambiguous identification of this phase in meteorites. SiC is likely to have formed in the atmospheres of late-type carbon-rich stars, where the conditions

C/O > 1 and Si/Mg > 1 were met¹³. The nucleosynthetic implications of the anomalous isotopic compositions are discussed in ref. 2.

We thank R. M. Walker for his constant interest and support for this research, and for his far-sighted effort which made possible the development of the microanalytical capabilities used in this work. This work was supported by grants from NSF (T.B., B.W. and E.Z.) and NASA (E.A. and T.M.)

Received 4 November; accepted 24 November 1987.

1. Tang, M. & Anders, E. *Geochim. cosmochim. Acta* (submitted).
2. Zinner, E., Tang, M. & Anders, E. *Nature* **330**, 730-732 (1987).
3. Feldman, D. W., Parker, J. H. Jr, Choyke, W. J. & Patrick, L. *Phys. Rev.* **170**, 698-704 (1968).
4. Feldman, D. W., Parker, J. H. Jr, Choyke, W. J. & Patrick, L. *Phys. Rev.* **173**, 787-793 (1968).
5. Krautwasser, P., Begun, G. M. & Angelini, P. *J. Am. ceram. Soc.* **66**, 424-434 (1983).
6. Lewis, R. S., Tang, M., Wacker, J. F., Anders, E. & Steel, E. *Nature* **326**, 160-162 (1987).
7. Wyckoff, R. W. G. *Crystal Structures*, 2nd edn, Vol. 1 (Wiley, New York, 1963).
8. Swart, P. K., Grady, M. M., Pillinger, C. T., Lewis, R. S. & Anders, E. *Science* **220**, 406-410 (1983).
9. Yang, J. & Epstein, S. *Nature* **311**, 544-547 (1984).
10. Halbout, J., Mayeda, T. K. & Clayton, R. N. *Earth planet. Sci. Lett.* **80**, 1-18 (1986).
11. Zinner, E. & Epstein, S. *Earth planet. Sci. Lett.* **84**, 359-368 (1987).
12. Wopenka, B. & Swan, P. D. *Meteoritics* **20**, 788-789 (1985).
13. Larimer, J. W. & Bartholomay, M. *Geochim. cosmochim. Acta* **43**, 1455-1466 (1979).

Large isotopic anomalies of Si, C, N and noble gases in interstellar silicon carbide from the Murray meteorite

Ernst Zinner*, Tang Ming† & Edward Anders†‡

* McDonnell Center for the Space Sciences and Department of Physics, Washington University, St Louis, Missouri 63130, USA

† Enrico Fermi Institute and Department of Chemistry, University of Chicago, Chicago, Illinois 60637, USA

Primitive meteorites contain several noble gas components with anomalous isotopic compositions which imply that they—and their solid 'carrier' phases—are of exotic, pre-solar origin. Having enriched two of these gas components by a factor of $\sim 2 \times 10^4$ in minor fractions of the Murray meteorite¹, we found that these fractions contain two minerals not previously seen in meteorites: silicon carbide and an amorphous Si-O phase², possibly an alteration product of another silicon-bearing mineral. Here we report ion microprobe analyses of these phases which reveal very large isotopic anomalies in silicon, nitrogen and carbon, exceeding the highest anomalies previously measured by factors of up to ~ 50 . We conclude that these phases are circumstellar grains from carbon-rich stars, whose chemical inertness allowed them to survive in exceptionally well-preserved form.

In a continuing search for the carriers of isotopically anomalous, pre-solar noble gas components, we treated a sample of the Murray C2 chondrite with HF-HCl to remove silicates and then with HClO₄ to remove organic and graphitic carbon¹. The residue (2,600 p.p.m.), consisting mainly of spinel (MgAl₂O₄) and diamond, was separated by grain size into five fractions. The coarsest and largest ($> 2 \mu\text{m}$, 1,860 p.p.m.) was treated first with H₃PO₄ to dissolve spinel and then with HNO₃ and HClO₄ to dissolve graphitic carbon released by the spinel, yielding residue CJ (30 p.p.m.). This sample and the second-finest fraction, CF (0.03-0.2 μm , 28 p.p.m.) had very high concentrations of Ne-E(H) (nearly pure ²²Ne) and Xe-S (enriched in the s-process isotopes 128, 130 and 132), respectively, exceeding the highest previously obtained values by factors of 9 and 20 (Table 1).

In absolute concentration, both samples are dominated by Ne-E, but CF, unlike CJ, also contains a substantial amount of Xe-S (Table 1). Both contain residual amounts of the very

abundant anomalous component Xe-HL (enriched in the heavy and light isotopes) and its carrier Cδ ($\sim 30\text{-}\text{\AA}$ diamond^{3,4}).

As reported in ref. 2, both samples contain crystalline SiC but CJ also contains an amorphous Si-O phase, possibly formed by alteration of a silicon-bearing mineral (nitride?) during the separation procedure. These silicon-bearing phases are the prime candidates for carriers of Ne-E(H) and Xe-S, as the other minerals in these samples are either gas-poor (spinel¹, chromite) or carry a different gas component (Cδ diamond).

For ion microprobe analysis, CJ was oxidized in an oxygen plasma asher for 30 min, to remove Cδ and most other elemental carbon, yielding residue CJa. The isotope measurements were made on a Cameca IMS 3F ion microprobe. Agglomerates of the residues were mounted on gold foil⁵ and carbon, nitrogen and silicon isotopes were measured on 5-10- μm spots as negative secondary ions^{5,6} produced by Cs⁺ bombardment. Nitrogen was measured as CN⁻ at a mass resolving power (MRP) of 6,500, high enough to separate ¹³C¹⁴N⁻ from ¹²C¹⁵N⁻ and ¹¹B¹⁶O⁻. During the nitrogen measurements, carbon was measured as C₂⁻, and during the silicon measurements as C⁻ (except for two spots where all three elements were analysed), at an MRP of 3,500. In many runs only the carbon isotopes and the Si⁻/C⁻ ratios were measured.

The standards used were NBS-21 graphite, synthetic SiC and 1-hydroxy-benzotriazole hydrate. All isotopic compositions are normalized for instrumental mass fractionation and are given as δ -values (see Fig. 1 legend) relative to the usual terrestrial standards (PDB, air and NBS-28 quartz for carbon, nitrogen and silicon, respectively). The errors represent the variations in the isotopic ratios during measurement, which were often much larger than the analytical precision, reflecting real heterogeneity in the grain aggregates. The Si/C ratios were obtained by normalizing the Si⁻/C⁻ ratios to that of the SiC standard.

The silicon in both samples is isotopically highly anomalous, as shown by comparison with the previous meteoritic range⁷

Table 1 Compositions of Murray samples

	CF	CJ
²² Ne-E(H) (10^{-8} ml g ⁻¹)	368	1,249
¹³⁰ Xe-S (10^{-10} ml g ⁻¹)	167	17
¹³⁶ Xe-HL (10^{-10} ml g ⁻¹)	17	69
SiC, crystalline (per cent)	~ 8	12
Si-O, amorphous (per cent)	~ 0	3
Carbon (per cent)	2	8
Spinel (per cent)	80	~ 5
Chromite (per cent)		~ 65

‡ Present address: Physikalisches Institut, University of Berne, CH-3012, Berne, Switzerland.

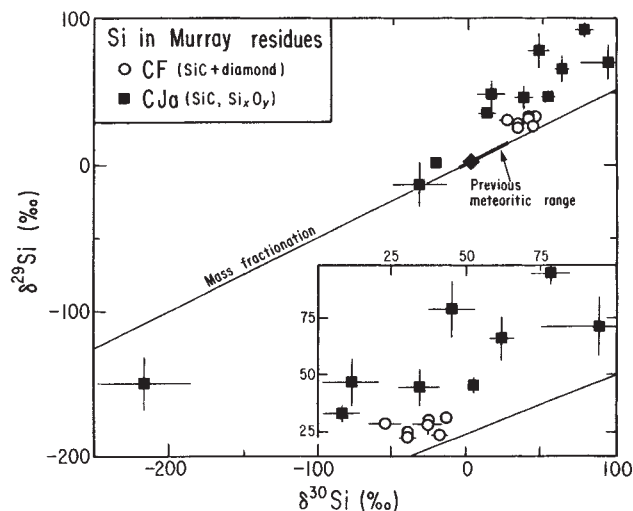


Fig. 1 $\delta^{29}\text{Si}$ plotted against $\delta^{30}\text{Si}$ for two highly processed, acid-resistant residues from the Murray meteorite. The residues have extremely variable silicon isotope compositions, clearly distinct from the terrestrial composition (diamond at 0,0) or previous meteoritic range⁷ (heavy line segment). The CF points cluster rather closely and lie only slightly above the mass fractionation line, but the CJa points show much greater variation. At least three components seem to be required: one high in ^{28}Si , one low in ^{28}Si and one high in ^{29}Si . Error bars represent sample heterogeneity, which is much larger than the analytical precision; where not shown, they are smaller than the size of the symbols. The inset more clearly shows the compositional variability, especially for the CF cluster. Data on C, N or both are available for a few points in the inset: the CJ point furthest to the left has $\delta^{13}\text{C} = 5,740\text{‰}$, $\text{Si/C} = 1.16$; the third CF point from the right has $\delta^{13}\text{C} = 770\text{‰}$, $\delta^{15}\text{N} = -375\text{‰}$, $\text{Si/C} = 0.64$. $\delta^{29}\text{Si} \equiv 1,000((^{29}\text{Si}/^{28}\text{Si})_{\text{sample}}/(^{29}\text{Si}/^{28}\text{Si})_{\text{standard}})$; other δ -values are defined analogously.

(heavy line in Fig. 1). Not only is the range greater, but all except two analyses lie above the terrestrial mass fractionation line, showing that these compositions cannot be derived from terrestrial silicon by mass fractionation. Indeed, as SiC is the major form of silicon under conditions in which it is stable⁸, little or no mass fractionation would be expected, and we shall therefore use the mass fractionation line only as a positional reference.

Previous data occasionally fell up to 1% above the line⁷, indicating a minor anomalous component, but the present data stray much further from the line. The data for CF (Fig. 1 and inset) cluster tightly at $\delta^{30}\text{Si} = 33\text{--}45\text{‰}$ and lie up to $11 \pm 3\text{‰}$ above the line, but the data for CJa show a much greater spread, $\delta^{30}\text{Si} = -217$ to $+95\text{‰}$, and lie up to $52 \pm 5\text{‰}$ above the line. This cannot be due to the higher areal density of CF, which provided a larger number of grains in the area of the beam and hence a smoother average composition: all CJa points above $\delta^{30}\text{Si} = 0$ fall consistently further from the line. At least part of this greater spread may be due to the Si-O phase, which is present in CJ but is rare or absent in CF. But even the SiC in the two samples may not be the same, as some of that in CJ originally may have been enclosed in spinel whereas that in CF was not.

The two extreme CJa points (farthest upper-right and farthest lower-left in Fig. 1) are collinear with the CF cluster and—perhaps significantly—with the terrestrial point. In principle, this set of points could be a binary mixture of two components lying at or beyond the extreme CJa points, one high and one low in ^{28}Si (with smaller differences in the other isotopes); however, nearly all of the remaining CJa points lie consistently above this putative mixing line, and require a third component with excess ^{29}Si .

The carbon isotope data are shown in Fig. 2. The Si/C ratios of CF are mainly <1 , reflecting the presence of C δ and other

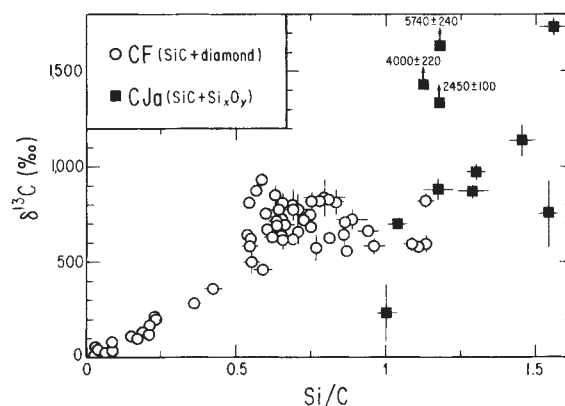


Fig. 2 $\delta^{13}\text{C}$ plotted against Si/C ratio for residues CF and CJa. The Si/C ratios of CF tend to be <1 owing to the presence of (isotopically light) diamond, whereas the ratios of CJa are higher, reflecting the presence of the Si-O phase. Most CF points lie between $\delta^{13}\text{C} = 600$ and 900‰ , but mixing lines from CS at -38‰ extrapolate to limits of 600 and $1,500\text{‰}$ at $\text{Si/C} = 1$, which may correspond to the principal type of SiC. Three of the CJa points (arrowed) lie substantially higher, off the scale, indicating the presence of at least one other SiC component, with $\delta^{13}\text{C}$ as high as $\sim 6,000\text{‰}$.

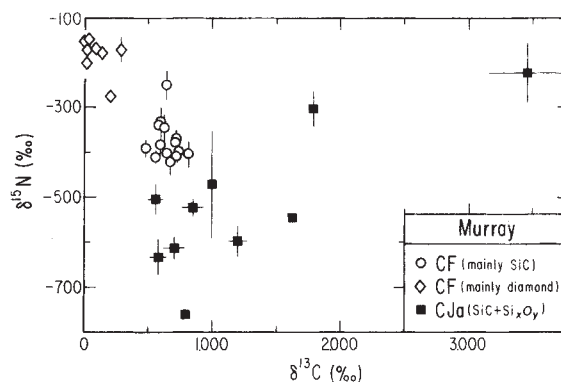


Fig. 3 $\delta^{15}\text{N}$ plotted against $\delta^{13}\text{C}$ for residues CF (grouped by dominant phases in the measured agglomerates) and CJa. CF apparently contains only a single, somewhat heterogeneous SiC component, of $\delta^{13}\text{C} = 600\text{--}900\text{‰}$ and $\delta^{15}\text{N} = -300$ to -400‰ . CJa contains a similar component, with virtually identical carbon but lighter nitrogen ($<-500\text{‰}$, as compared with -300 to -400‰), and a second component with heavier carbon and nitrogen, at or beyond the point in the upper right-hand corner.

elemental carbon (possibly from spinel⁶), whereas those of CJa are higher, owing to removal of much of the carbon and the presence of the Si-O phase.

Whereas there is little spread in the $\delta^{13}\text{C}$ of the CF data points at low Si/C ratios, suggesting only a single pure carbon component (C δ diamonds, with $\delta^{13}\text{C} = -38\text{‰}$ ⁹), the $\delta^{13}\text{C}$ values at higher Si/C ratios scatter more widely, and extrapolate to 600– $1,500\text{‰}$ at $\text{Si/C} = 1$. This is also the range of most CJa points, indicating similar, if somewhat heterogeneous, populations in both separates. (The Si-O phase, containing only minor carbon, contributes very little to the $\delta^{13}\text{C}$ values, although it increases the Si/C ratio.) But three of the eleven CJa points plot substantially higher, to $+6,000\text{‰}$. This again implies that at least one other SiC component, of higher $\delta^{13}\text{C}$, is present in CJa.

The nitrogen data are shown in Fig. 3. Because nitrogen is measured as CN^- , it comes mainly from SiC, although a small amount may be contributed by a minor, unidentified nitrogen-bearing phase². The carbon data, on the other hand, are dominated by SiC, as before.

The CF points with Si/C < 0.5 (open diamonds), which have large contributions from C δ diamonds, form a distinct cluster at $\delta^{15}\text{N} = -150$ to -200% , not quite as light as expected for C δ (-375% , according to a previous bulk analysis¹⁰). The central CF-CJ band from Fig. 2 ($\delta^{13}\text{C} = 600$ – 900%) has broken up; the CF points cluster between $\delta^{15}\text{N} = -300$ and -400% , again suggesting a single, somewhat heterogeneous component, but the CJa points fall consistently lower, down to $\delta^{15}\text{N} = -762\%$. Three possibilities come to mind which will have to be checked in future work: (1) CF and CJa contain different populations of SiC; (2) CF contains a combustible phase with heavier nitrogen, which raises the SiC- and C δ -rich points; and (3) CJa has variable contributions of very light nitrogen from either the Si-O phase or some undetected Si-N mineral, as the CJa points often have up to fourfold higher CN⁻/C₂⁻ ratios². Interestingly, two of the three CJa points with high $\delta^{13}\text{C}$ have some of the least negative $\delta^{15}\text{N}$ values in the sample (-303 and -222%), and since these values presumably still have contributions from the abundant light nitrogen in this sample, the true $\delta^{15}\text{N}$ may well be much higher. Such a composition would be expected for the carrier of Ne-E(H), which presumably came from a nova¹¹, as novae produce large amounts of ^{13}C and ^{15}N .

The principal results and implications of this work are as follows:

(1) Silicon carbide is apparently the phase responsible for the isotopically heavy carbon that has been seen in C1-C2 chondrites on stepped combustion at $\geq 900^\circ\text{C}$ ($\delta^{13}\text{C} = 1,100$ – $1,500\%$ ^{9,12}). The high combustion temperature pointed to a phase less combustible than diamond¹³, and SiC, which oxidizes slowly even at $1,000^\circ\text{C}$, meets this requirement. At least two kinds of SiC seem to be present: a possibly complex one with $\delta^{13}\text{C} \approx 600$ – $1,500\%$ and $\delta^{15}\text{N} \approx -300$ to -400% (CF) or even -760% (CJ, but perhaps the differences reflect the presence of another nitrogen component in one or the other sample), and a second kind in CJ with heavier carbon and nitrogen (to $+6,000\%$ and at least -222% ; respectively).

(2) Both noble gas components are presumably located in SiC, but as the SiC components are not well resolved, only tentative assignments can be made. Xenon-S is strongly concentrated in CF (Table 1) and thus may be located in its principal type of SiC. Neon-E(H), being present mainly in CJ, could be located in one of the types of SiC in that sample. The Si-O phase, if indeed secondary, is not likely to have retained neon. A more definite assignment will require purer samples.

(3) Presumably the several types of SiC are circumstellar condensates from highly evolved stars with C/O > 1, as predicted by thermodynamic calculations^{8,14,15} and confirmed by astronomical observations¹⁶. Nothing definite can yet be said about the unidentified precursor (silicon nitride or oxynitride?) of the Si-O phase, although such phases, too, can presumably form as primary or secondary condensates at high C/O ratios¹⁵.

(4) The heavy carbon and light nitrogen of the major SiC component are qualitatively consistent with production in the CNO cycle. Silicon, on the other hand, is not affected by hydrogen burning, and so the silicon anomalies must have been produced in later stages of nucleosynthesis. At least three components are required: one high in ^{28}Si , one low in ^{28}Si and one high in ^{29}Si . Xe-S requires the s-process, whereas Ne-E—if derived from ^{22}Na , with a 2.6-yr half life—requires an explosive process, for example in a nova. The discrete nature of the various isotopic components suggests several individual stars, but a single, massive supernova is an interesting alternative¹⁷. Such a supernova can at least qualitatively account for the observed anomalies¹⁷, and although it would give a continuum of compositions, chemical and other factors may cause selective sampling of this continuum.

(5) There are two indications that the SiC grains are relatively young, and were incorporated in the meteorite soon after their formation. First, ^{21}Ne in CJ gives a cosmic-ray exposure age of < 60 Myr for the SiC¹⁸, which is less than the estimated 10^9 -yr

lifetime of typical interstellar grains¹⁹. Second, the size distribution of the SiC in a sister sample of CF is log-normal and sharply peaked¹, as expected for a primary crystal growth distribution that has not been altered by erosion or fragmentation.

(6) It is curious that the isotopic anomalies for silicon, carbon, nitrogen and associated noble gases are 10–1,000 times larger than those for most other elements in meteorites. One reason is the resistance of SiC and carbon to alteration by heat or liquid water in the meteorite parent bodies, in contrast to the major meteoritic minerals. A more profound reason is the low C/O ratio (~ 0.6) of the Solar System and hence of its principal stellar sources²⁰, which precludes condensation of phases such as SiC or carbon that require C/O ratios $> 0.83^8$. These phases thus must come from carbon-rich sources that made only a minor contribution to the Solar System (note, in this connection, that the silicon in SiC constitutes only $\sim 2 \times 10^{-5}$ of the total silicon in the meteorite). It is fortunate that these exotic phases come to us undiluted by local counterparts, and in a chemically resistant form that permits them to survive destruction processes in interstellar space, meteorite parent bodies and—with the exception of the precursor of the Si-O phase—the laboratory.

Received 2 November; accepted 27 November 1987.

1. Tang, M. & Anders, E. *Geochim. cosmochim. Acta* (submitted).
2. Bernatowicz, T. et al. *Nature* **330**, 728–730 (1987).
3. Lewis, R. S., Tang, M., Wacker, J. F., Anders, E. & Steel, E. *Nature* **326**, 160–162 (1987).
4. Lewis, R. S., Bright, D. & Steel, E. *Meteoritics* **22** (in the press).
5. McKeegan, K. D., Walker, R. M. & Zinner, E. *Geochim. cosmochim. Acta* **49**, 1971–1987 (1985).
6. Zinner, E. & Epstein, S. *Earth planet. Sci. Lett.* **84**, 359–367 (1987).
7. Clayton, R. N., Mayeda, T. K. & Molini-Velsko, C. A. in *Protostars and Planets II* (eds Black, D. C. & Matthews, M. S.) 755–771 (University of Arizona Press, Tucson, 1985).
8. Lattimer, J. M., Schramm, D. N. & Grossman, L. *Astrophys. J.* **219**, 230–249 (1978).
9. Swart, P. K., Grady, M. M., Pillinger, C. T., Lewis, R. S. & Anders, E. *Science* **220**, 406–410 (1983).
10. Ash, R. D. et al. *Meteoritics* **22** (in the press).
11. Clayton, D. D. & Hoyle, F. *Astrophys. J.* **203**, 490–496 (1976).
12. Yang, J. & Epstein, S. *Nature* **311**, 544–547 (1984).
13. Pillinger, C. T. *Eur. Space Ag. spec. Publ. SP-249*, 41–45 (1986).
14. Gilman, R. C. *Astrophys. J.* **155**, L185–L187 (1969).
15. Larimer, J. W. & Bartholomay, M. *Geochim. cosmochim. Acta* **43**, 1455–1466 (1979).
16. Kleinmann, S. G., Gillett, F. C. & Joyce, R. R. *Astr. Astrophys.* **19**, 411 (1981).
17. Arnett, D. *Nature* (submitted).
18. Tang, M. & Anders, E. *Geochim. cosmochim. Acta* (submitted).
19. Salpeter, E. E. *A. Rev. Astr. Astrophys.* **15**, 267–293 (1977).
20. Schramm, D. N. & Olive, K. A. *Ann. N.Y. Acad. Sci.* **395**, 236–241 (1982).

Simultaneous observation of fundamental and second harmonic radio emission from the terrestrial foreshock

D. Burgess*, C. C. Harvey†, J.-L. Steinberg† & C. Lacombe†

* Mullard Space Science Laboratory, Dorking, Surrey RH5 6NT, UK
† DESPA, Observatoire de Paris-Meudon, 92195 Meudon, France

There have been many reports of the terrestrial foreshock as a source of freely propagating electromagnetic (e.m.) radiation^{1–7}, identified as plasma emission at twice the solar-wind plasma frequency, $2f_p$. The $2f_p$ source is close behind the curved surface defined by the interplanetary magnetic field lines tangent to the Earth's bow shock. Electrons accelerated at the shock over a wide range of energies propagate along these field lines where, from a time-of-flight effect⁶, they have a distribution that is unstable to the production of Langmuir waves^{6,8}. These Langmuir waves^{9,10} and the unstable electron distributions^{11,12} have been observed. The mechanism suggested for generating the $2f_p$ radiation^{5,6,13,14} by the terrestrial foreshock is closely related to present models of type III solar radio emission and rely on some of the energy in the Langmuir waves converting to e.m. waves. The e.m. emission

of type III bursts is known to be produced in the solar wind at both the fundamental and second harmonic¹⁵, and whether fundamental emission is produced in the terrestrial foreshock is of considerable interest. Comparison of nearby fundamental and harmonic emissions and the Langmuir wave energy density in the source will help discriminate between the different conversion models^{8,14,16}. We present here unambiguous measurements made with the sounder experiment aboard the International Sun-Earth Explorer-1 (ISEE-1) and the radio experiment on ISEE-3 of simultaneous fundamental and second harmonic e.m. emission from the Earth's foreshock.

For a temporally and spatially near uniform solar-wind density it is not feasible, with an electric antenna, to measure reliably e.m. radiation close to f_p . Below f_p , e.m. waves are evanescent; at and above f_p any weak e.m. component would be masked either by the natural quasi-thermal noise^{7,17}, or, within the foreshock, by intense electron plasma waves^{9,10}. These comments depend, of course, on the bandwidth and shift of the fundamental emission; but, according to the electron beam model¹⁴, the relative bandwidth of the fundamental cannot be larger than the relative bandwidth of the $2f_p$ radiation which, we note, is fairly small⁵ ($\Delta f/f \approx 0.01$). Any turbulence present would tend to scatter and diffuse the radiation. Thus, in order to see fundamental emission from the foreshock there must be a low density path from the emitting region to the receiver, that is, a drop in the solar-wind density at the receiver relative to the density at some part of the foreshock.

From 1981 day 242 (30 August 1981) 20:00 UT to day 243 (31 August 1981) 21:00 UT ISEE-1 and -2 were in the solar wind and the propagation and sounder experiments were continuously active in high-bit-rate mode. The propagation experiment¹⁸ determines the integrated electron density between the two spacecraft by measuring the phase advance introduced onto a wave at 683 kHz 32 times per second. The sounder¹⁸ is essentially a sweep frequency analyser with 128 steps of bandwidth 400 Hz and separation 395 Hz, taking 16 s to complete a sweep from 0.5 to 51 kHz. When active (as here) at each frequency step a short pulse is emitted at the centre frequency. This can excite various plasma resonances which are detected by the receiver subsequently taking 7 consecutive samples at the same frequency. The instrument then steps to the next frequency, each step lasting 0.125 s. The plasma resonances can be identified from their characteristic time decay.

For the period from day 242 at 21:00 UT to day 243 at 12:00 UT the solar wind was unusual in that the density made jumps of considerable magnitude separated by some periods of relative calm, and others of considerable turbulence. As the density changes were resolved by both experiments, they are not fully developed magnetohydrodynamic discontinuities, which would be possible sources of accelerated particles and turbulence. Figure 1 shows a spectrogram of sounder data from 06:29 to 07:07 UT, when the geocentric solar ecliptic coordinates (GSE) of ISEE-1 were (19.6, 11.2, -3.1) Earth radii (R_E). The spacecraft separation was only 50 km, so that the propagation experiment gives a good measure of the local density. The sounder resonance at the local plasma frequency is evident after 06:38, and follows the propagation density. Between 06:30:45 and about 06:38 the plasma resonance is too close to the 40-kHz interference line to be seen clearly. From the propagation data it is clear that between 06:30:30 and 06:50:00 a region of enhanced solar-wind density passed over the spacecraft.

Between 06:48 and 06:50:30 the solar-wind density decreases from 20 to 12 cm^{-3} . The magnetic field remains approximately constant in magnitude and direction, being roughly perpendicular to the ecliptic plane, northwards. Most unusual is the emission that can be seen after the trailing edge of the density pulse. There is a line feature centred at ~ 37 kHz which appears as the continuation of the local plasma frequency seen earlier between 06:43 and 06:48. This continues with some variations in intensity and centre frequency until 07:04 when it fades away.

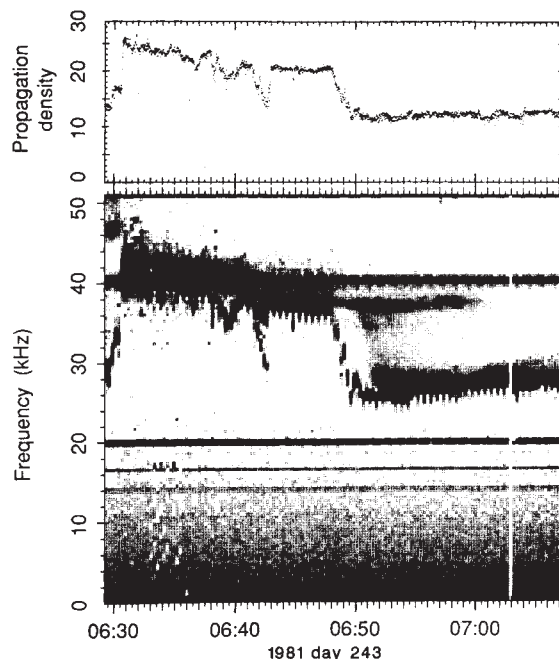


Fig. 1 The top panel shows the density determined by the propagation experiment. Below, a time-frequency spectrogram of data from the ISEE-1 sounder, showing average power on a logarithmic scale running from $10^{-15.5}$ (white) to $10^{-13.4} \text{ V}^2 \text{ m}^{-2} \text{ Hz}^{-1}$ (black). The continuous operation of the propagation experiment introduces interference lines at 20 and 40 kHz, and two slightly weaker ones at 14 and 17 kHz.

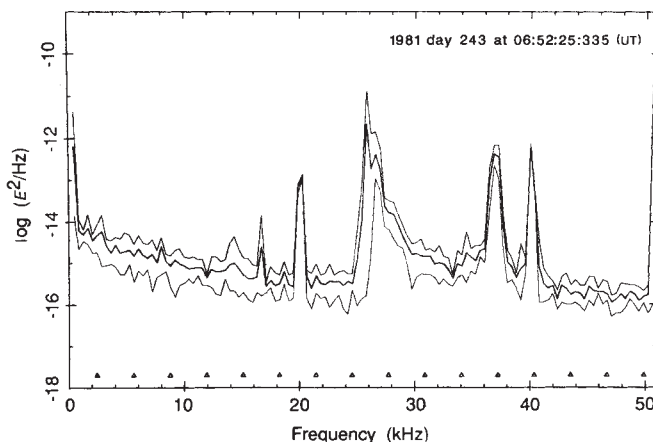


Fig. 2 A spectrum from the ISEE-1 sounder taken from during the period shown in Fig. 1. Shown are the minimum, maximum and average of the seven electric field power measurements on each frequency step. The fluctuations are statistical, corresponding to noise measured with a receiver of integration time of 2 ms, except at 25 kHz where they are due to the decay of the plasma resonance.

During this emission the local plasma frequency resonance is ~ 26 kHz, so that the 37-kHz emission is evidently not $2f_p$ radiation. It is our contention that the 37-kHz radiation is fundamental e.m. emission from a region of the foreshock where the solar-wind density has not yet changed from the value measured at ISEE-1 between 06:43 and 06:48. Thus, the emission provides a remote sensing of the density somewhere at the foreshock, although we cannot tell which part with the sounder, because it is ill-adapted measure spin modulation. We can only conclude that for the duration of the emission the solar-wind density was strongly inhomogeneous on a scale smaller than the foreshock. This is corroborated by the fact that the density change starting

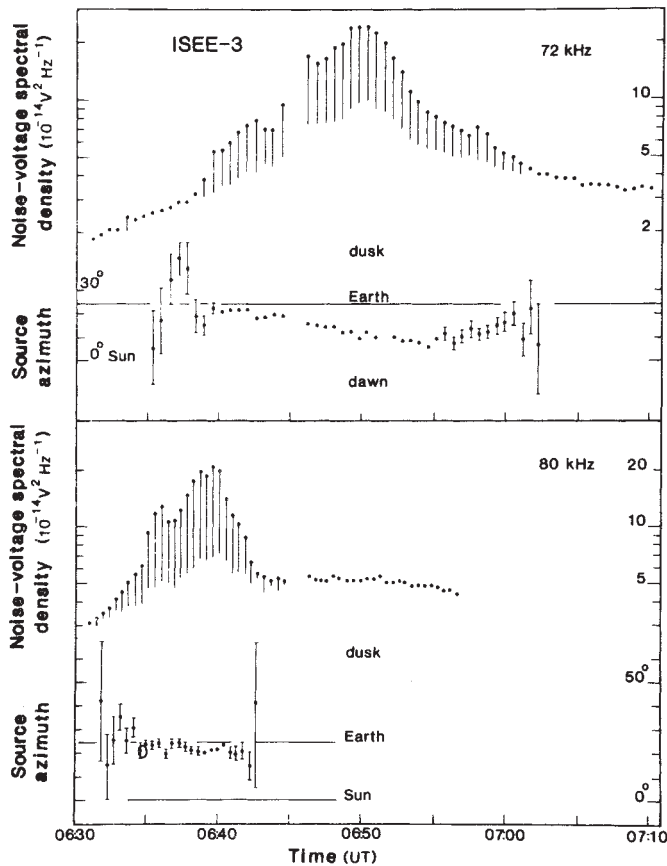


Fig. 3 Noise-voltage spectral density from two channels of the ISEE-3 radioastronomy experiment. The antenna electrical length is 45 m, so that the maximum electric field power at 72 and 80 kHz is $\sim 10^{-17} \text{ V}^2 \text{ m}^{-2} \text{ Hz}^{-1}$. The peak-to-peak amplitude of the spin modulation is shown by vertical lines, and the source azimuth indicates that the emission seen by ISEE-3 at 70–80 kHz is terrestrial.

at 06:48 lasted only 2.5 min, with a corresponding convected distance of $10 R_E$ (taking $V_{sw} = 400 \text{ km s}^{-1}$ where V_{sw} is the solar-wind velocity), much smaller than the terrestrial $2f_p$ source ($100 R_E$)⁵.

In order to separate resonance features from naturally occurring features we show in Fig. 2 the spectral sweep starting at 06:52:25. The maximum, minimum and average electric field power show that the 37-kHz emission is fairly constant, with little variation during each frequency step, in contrast to the electrostatic noise observed in the foreshock, which shows extremely rapid variations of amplitude^{9,10}. A natural explanation is that the emission is e.m. radiation which has propagated from some distant source. Unlike the plasma resonance at 25 kHz, the 37-kHz emission does not show any decay of amplitude over a frequency step, which indicates that it is not an artefact of the sounder activity. This emission is fairly intense ($E_r^2 \approx 10^{-12.5} \text{ V}^2 \text{ m}^{-2} \text{ Hz}^{-1}$), but it is difficult to draw from this any quantitative conclusion about the source intensity because of our uncertain knowledge of the plasma and magnetic field along the path to the observer, and the possibility of source anisotropy. But assuming a source subtending a solid angle of 2π , we can estimate the brightness temperature of the 37-kHz fundamental emission to be approximately $T_b = 10^{15} \text{ K}$.

Returning to Fig. 1, we point out that from 06:51 to 06:57 there is also continuum emission between the sounder resonance and the 37-kHz emission. At any fixed frequency there is a constant delay of ~ 2 min from the drop in the local plasma frequency to the start of the continuum. This is exactly the

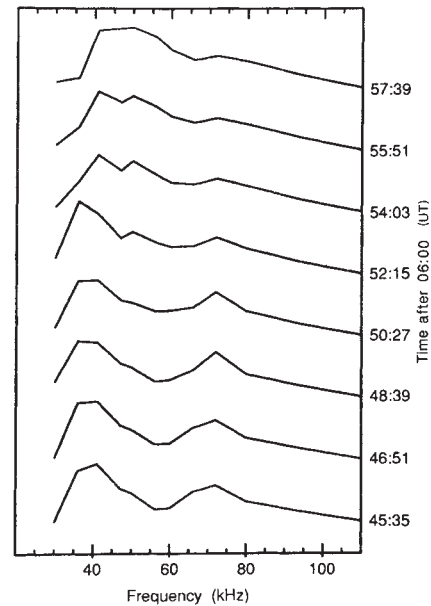


Fig. 4 A time series of spectra, in arbitrary flux units, each averaged over 108 s, from the ISEE-3 receiver. The time of each spectrum after 06:00 UT is marked alongside. Notice the feature that appears after 06:52:15 at 50 kHz, corresponding to the second harmonic of the local plasma frequency at ISEE-1. The emission at 72 kHz corresponds to that shown in Fig. 3.

behaviour expected if we suppose that the density change seen at 06:48 will emit fundamental radiation once the trailing edge of the density pulse reaches the foreshock. The two minute delay corresponds to a convected distance of $8 R_E$. This is not inconsistent with the position of ISEE-1 and the position of the foreshock (effectively a curved sheet perpendicular to the ecliptic plane, passing through the nose of the bow shock, nominally at $X_{GSE} = 14 R_E$). Several other instances of the same behaviour, and the same delay, are observed on the same day. The exact relationship between the observed delay and the solar-wind speed should take into account the orientation of the density change; the position of ISEE-1 (at 16:00 local solar time) indicates that the normal of the interplanetary gradient is directed towards the dusk side of the Earth.

Comparing foreshock and type III e.m. f_p emissions, we note that both are observable thanks to a solar-wind density gradient which, however, is much steeper and more localized in the foreshock case. Both continua exhibit a leading edge which falls in frequency as time progresses; in the foreshock this is due to the density gradient convecting across an electron beam, rather than an electron beam propagating into the density gradient.

If indeed the emission at 37 kHz from 06:38 to 07:04 is fundamental and generated at the foreshock, and admitting that the second harmonic radiation is generated in the same region, then we should expect to see the second harmonic at 74 kHz. As the ISEE-1 receiver went up to only 51 kHz, we turn to observations made by the radio astronomy experiment¹⁹ aboard ISEE-3. The receiver has neither the temporal nor frequency resolution of the ISEE-1 sounder, but is sensitive and well adapted to measuring the spin modulation of a signal and hence, if it is localized, its azimuth direction in the ecliptic plane. During this period ISEE-3 was at $(238.8, -94.2, 10.5) R_E$ (GSE), so that the azimuth angle between the anti-Sun direction and the Earth direction was $\sim 23^\circ$.

For the period of interest the ISEE-3 data shows a type III burst extending into the frequencies where we must look for a terrestrial $2f_p$ signal. We can, however, disentangle the type III burst and any $2f_p$ radiation by examining the spin modulation; terrestrial $2f_p$ radiation is generally narrow-band, well modu-

lated and in the direction of the Earth^{3,5}. We first identify the ISEE-3 receiver channels where we expect to see $2f_p$ radiation. From 06:31 to 06:39 the ISEE-1 local f_p is ~ 40 kHz, and so $2f_p$ would be in the ISEE-3 80(3)-kHz channel (the number in brackets gives the bandwidth in kHz of the channel). Until 06:48 the local plasma frequency is 37 kHz and the corresponding ISEE-3 channel is 72(10)-kHz. Between 06:48 and 07:03 ISEE-1 sees radiation at 37 kHz, which we denote by f_p^* and, as argued above, we expect $2f_p^*$ radiation to be seen simultaneously at ISEE-3.

In Fig. 3 we show the noise-voltage spectral density, spin-modulation, and signal-azimuth direction for the channels of interest. In the 80-kHz channel there is a spin-modulated, Earth-directed signal from 06:32 to 06:44, and at 72 kHz a similar signal from 06:39 to 07:02. If these two signals are second harmonic radiation of the emission observed by ISEE-1, then their partial time overlap simply reflects the bandwidths of the channels and the drift of the local plasma frequency from 06:32 to 06:48 (Fig. 1). Higher frequency channels do not show spin modulation.

The most important point from Fig. 3 is that there is a $2f_p^*$ signal in the 72-kHz channel from 06:48 to 07:02, exactly when at ISEE-1 we observe the emission we have identified as fundamental. This is strong evidence that the emission f_p^* observed at ISEE-1 is produced in the same region as the $2f_p^*$ observed at ISEE-3, namely the terrestrial foreshock. It is interesting to note that the direction of the $2f_p^*$ radiation (for example, at 06:55) is different from the $2f_p$ radiation observed in the 80-kHz channel at 06:40. This behaviour of the apparent source azimuth must be a consequence of the source distribution on the foreshock, the orientation of the interplanetary density gradient and any intrinsic source directivity.

It has been crucial to our argument that for the period 06:48 to 07:04 there are two spatially distinct values for the solar-wind density at the foreshock, so we should expect to see at ISEE-3 a further emission, at twice the local ISEE-1 plasma frequency. Between 06:50 and 07:00 the sounder resonance is ~ 26 kHz, and any $2f_p$ signal at ISEE-3 should be seen in the 50(10) kHz channel. In Fig. 4 we show a time sequence of spectra measured at ISEE-3 between 06:45:35 and 06:57:39, each spectrum being an average over 108 s. The line feature at 72 kHz discussed above is clearly visible. There is a broader, less prominent continuum-like feature starting at ~ 60 kHz, and extending to higher frequencies, which is the tail of the type III burst. There is a further contribution in the 50-kHz channel which first appears in the spectrum of 06:52:15, and which coincides with the drop of the local plasma frequency at ISEE-1 at 06:50, suggesting that it is second harmonic emission corresponding to the local f_p at ISEE-1. It is difficult to follow the evolution of this feature because of the type III emission, and the fact that after about 06:52 the local plasma frequency at ISEE-3 (seen as the lower frequency peak and cutoff in Fig. 4) rises to ~ 40 kHz (confirmed by the ISEE-3 electron experiment). No spin modulation is seen in the 50-kHz channel, but this is not surprising because near the local plasma frequency one expects radiation to be strongly diffused by density irregularities. Although not prominent in Fig. 4, the 50-kHz signal is well above the quasi-thermal noise¹⁷ as predicted from the electron parameters measured simultaneously at ISEE-3. Thus, although not conclusive, the evidence points towards the signal in the 50-kHz channel being the second harmonic of the plasma frequency measured locally by ISEE-1 after 06:50.

The present observations further strengthen the analogy between type III solar radio bursts and the processes occurring in the terrestrial electron foreshock. The numerous *in situ* observations of the foreshock may throw light upon the processes which occur in type III radiobursts: initial beam instability, and subsequent stabilization through nonlinear wave-wave and wave-particle interactions. These are processes which must occur quite generally in plasma astrophysics.

One of the authors (D.B.) acknowledges a Royal Society European Exchange Fellowship, funded by the CNRS. We thank R. Zwickl of the Los Alamos National Laboratory for providing electron parameters from the ISEE-3 solar-wind experiment.

Received 10 June; accepted 10 October 1987.

1. Duncel, N. thesis, Stanford Univ. (1974).
2. Gurnett, D. A. *J. geophys. Res.* **80**, 2751-2763 (1975).
3. Hoang, S., Fainberg, J. & Steinberg, J.-L. *J. geophys. Res.* **86**, 4531-4536 (1981).
4. Harvey, C. C., Etcheto, J. & Mangeney, A. *Space Sci. Rev.* **23**, 39-58 (1979).
5. Lacombe, C., Harvey, C. C., Hoang, S., Mangeney, A., Steinberg, J.-L. & Burgess, D. *Ann. Geophys.* (in the press).
6. Filbert, P. C. & Kellogg, P. J. *J. geophys. Res.* **84**, 1369-1381 (1979).
7. Steinberg, J.-L. & Hoang, S. *Ann. Geophys.* **4A**, 429-438 (1986).
8. Cairns, I. H. *J. geophys. Res.* **92**, 2329-2342 (1987).
9. Lacombe, C., Mangeney, A., Harvey, C. C. & Scudder, J. D. *J. geophys. Res.* **90**, 73-94 (1985).
10. Etcheto, J. & Fauchoux, M. *J. geophys. Res.* **89**, 6631-6653 (1984).
11. Anderson, R. R., Parks, G. K., Eastman, T. E., Gurnett, D. A. & Frank, L. A. *J. geophys. Res.* **86**, 4493-4510 (1981).
12. Fitzenreiter, R. J., Klimas, A. J. & Scudder, J. D. *Geophys. Res. Lett.* **11**, 496-499 (1984).
13. Cairns, I. H. & Melrose, D. B. *J. geophys. Res.* **90**, 6637-6640 (1985).
14. Cairns, I. H. *J. geophys. Res.* (submitted).
15. Dulk, G. A., Steinberg, J.-L. & Hoang, S. *Astr. Astrophys.* **141**, 30-38 (1984).
16. Goldman, M. V. *Sol. Phys.* **89**, 403-442 (1983).
17. Couturier, P., Hoang, S., Meyer-Vernet, N. & Steinberg, J.-L. *J. geophys. Res.* **86**, 11127-11138 (1981).
18. Harvey, C. C., Etcheto, J., de Javel, Y., Manning, R. & Petit, M. *IEEE Trans. Geosci. Electr.* **GE-16**, 231-242 (1978).
19. Knoll, R. *et al. IEEE Trans. Geosci. Electr.* **GE-16**, 199-204 (1978).

Optical phase obtained by analogue Hartley transformation

John Villasenor & R. N. Bracewell

Department of Electrical Engineering, Stanford University, Stanford, California 94305, USA

When the two-dimensional Fourier transformation is performed with a lens the optical amplitude and phase in the output plane represent the complex transform. It can be shown that the two-dimensional Hartley transform is mathematically equivalent to the Fourier transform, but is real valued; amplitude alone fully represents everything. This is significant because ordinary optical detectors do not respond to phase. Here we describe the construction of an optical system in the form of a modified Michelson interferometer which physically demonstrates that it is possible to produce the Hartley transform of a plane luminous object. It is thus possible to encode in the form of amplitude the half of the information in a diffraction pattern that normally is carried in the form of phase.

A transform proposed in 1942 by Hartley¹ lay dormant until computer advances gave it relevance to spectral analysis² and to numerical image processing³, where it possesses a speed advantage over the fast Fourier transform algorithm. Unlike the Fourier transform, which is complex, the Hartley transform is purely real but fully equivalent. This property is of interest to optics, where Fourier transformation is commonly practised by means of a suitable lens arrangement. In the Fourier transform plane the information resides both in the amplitude and phase of the optical excitation over the plane but the phase information cannot be captured by photographic or other ordinary light detectors. Thus if the Hartley transform could be constructed it would have significant application. The Hartley transform does not need to be complex because it does not possess rotational symmetry. The Fourier transform by contrast is hermitian, having twofold rotational antisymmetry; it packs all its information content on one half of the plane, requiring two parameters (amplitude/phase or real/imaginary) per point. The other half of the plane is occupied by redundant information. The Hartley transform, being non-redundant, carries only one parameter per point, in the form of amplitude, over the whole plane.

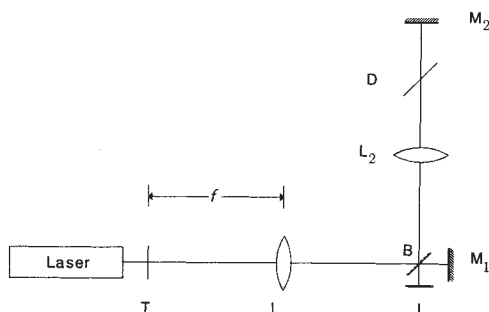


Fig. 1 Optical system for taking the Hartley transform of a transparency T in the (x, y) -plane using lenses L_1 and L_2 of focal length f , mirrors M_1 and M_2 , beam-splitter B , and delay D . The transform appears at plane I .

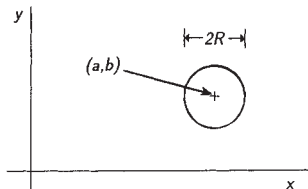


Fig. 2 A transparency T with transmittance zero except over a displaced disk of radius R centred at (a, b) .

Several schemes have been suggested⁴ for performing the Hartley transformation optically and we have chosen the simplest as an initial demonstration.

The two-dimensional Hartley transform $H(u, v)$ of a real function $f(x, y)$ is defined in terms of the cas function, that is, $\text{cas } \theta = \cos \theta + \sin \theta$ by

$$H(u, v) = \iint f(x, y) \text{cas}[(2\pi(ux + vy))] dx dy$$

Let

$$F(u, v) = \iint f(x, y) \exp[-i2\pi(ux + vy)] dx dy$$

be the Fourier transform of $f(x, y)$. From the Hartley transform one can recover the Fourier transform through

$$2F(u, v) = H(u, v) + H(-u, -v) - i[H(u, v) - H(-u, -v)].$$

The Fourier phase $\phi = \arctan(\text{Im}F/\text{Re}F)$ is obtainable directly from the Hartley transform without reference to the complex Fourier transform at all through $\tan(\phi + \pi/4) = H(-u, -v)/H(u, v)$. Since $F(-u, -v)$ is $F(u, v)$ rotated through π in the (u, v) -plane, the relation

$$H(u, v) = [F(u, v) + \exp(-i\pi/2)F(-u, -v)] \times \exp(i\pi/4)/\sqrt{2}$$

shows that the Hartley transform can be constructed by superimposing a Fourier transform $F(u, v)$ on a phase-quadrature rotated copy of itself. The unimportant factor $\exp(i\pi/4)/\sqrt{2}$ simply reflects conventions of time origin and amplitude that come from Fourier theory.

Figure 1 shows the modified Michelson interferometer system that we used to perform the Hartley transformation by implementing this superposition. The input transparency T with transmission factor $f(x, y)$ is placed in the back focal plane (x, y) of a lens L_1 and coherently illuminated from behind. The portion of the beam that reaches mirror M_1 is then redirected by the beamsplitter B to form a Fourier transform on an output plane I , the (u, v) -plane. A second beam forms a Fourier transform at an intermediate plane near B ; then lens L_2 produces a rotated copy of the object at mirror M_2 . After reflection from M_2 , the beam passes again through L_2 , through the beamsplitter and on to the image plane, where it contributes the rotated

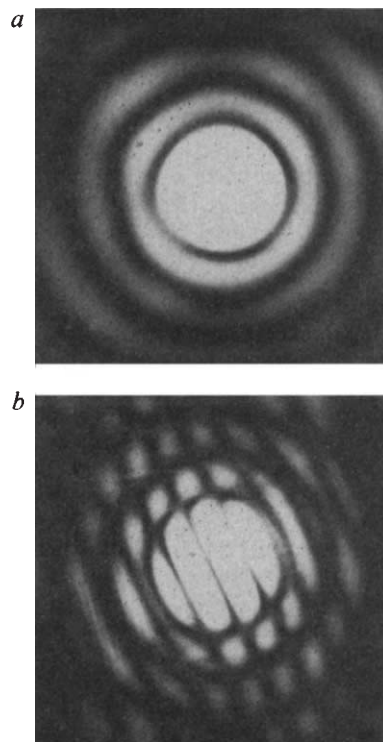


Fig. 3 *a*, Fourier intensity of the displaced object in Fig. 2 is independent of the displacement (a, b) . *b*, The Hartley intensity of the same object encodes the Fourier phase, lost in a photographic record of intensity, in the form of fringes whose orientation indicates the direction $\arctan(b/a)$ of displacement and whose period is inversely proportional to the displacement $d = \sqrt{a^2 + b^2}$.

Fourier transform $F(-u, -v)$. If the focal region were vanishingly thin it would be necessary for M_1 , B and I to all occupy the same point in space, an obvious impossibility in Fig. 1. The actual depth of the focal region is non-zero and depends on the lens focal length: diameter ratio f/D ; if this ratio is large the longitudinal dimension of the focal region becomes sufficient to permit displacement of M_1 and I . This can also be compensated for by slight displacement of lens L_2 , although this introduces a small quadratic phase error in the extremity of the transform. Other methods for obtaining the optical Hartley transform^{2,4} requiring slightly more complex equipment exist that do not include a beamsplitter or mirror near the output plane.

As there are two beams, the quadrature relation between $F(u, v)$ and $F(-u, -v)$ could be produced in principle without further optical elements by correctly positioning M_1 and M_2 . In practice a test transparency may be used to position M_1 (the second beam being interrupted) so that the expected Fourier transform is seen at I ; and similarly with M_2 . With both beams active, $F(u, v)$ and $F(-u, -v)$ will be superimposed in some arbitrary phase relation that has to be brought to $\frac{1}{2}\pi$. A variable delay D , such as a rotatable oblique glass plate in one arm, is suitable for phase adjustment.

The phase adjustment is correct, given a point source offset from the (x, y) origin by a distance d wavelengths in an arbitrary direction, when an appropriately oriented cas function, of angular period d^{-1} , appears at I . Such a pattern is distinguishable from sinusoidal fringe patterns in general because a fringe intensity maximum departs by one-quarter of a fringe spacing from the (u, v) origin.

Phase adjustment is not necessarily critical; it has already been shown⁵ that the kernels $\sqrt{2}\cos(\theta - \alpha)$, where $0 < \alpha < \frac{1}{2}\pi$, also lead to real, reciprocal transforms. In fact, the sum of $F(u, v)$ and $F(-u, -v)$ with any relative phase will yield a real

distribution, though $\alpha = \frac{1}{4}\pi$ is desirable in that the even and odd portions of the object are given equal weight in the transform. It is sufficient for α to be near $\frac{1}{4}\pi$ for the characteristic properties of the Hartley transform to be realized.

The quarter-fringe departure of the fringe system from the origin does not require absolute determination; the precise displacement makes itself apparent as soon as the offset point source has appreciable extent, as explained next.

The transparency (Fig. 2) is a circular aperture centred at (a, b) on an opaque field. A disc of radius R located at the origin has identical Hartley and Fourier transforms proportional to $\text{jinc } 2Rq$, where $\text{jinc } q$, the two-dimensional Fourier transform of a circular pupil of unit diameter, is given by $\text{jinc } q = J_1(\pi q)/2q$, and $q = \sqrt{u^2 + v^2}$ is the radial coordinate in the (u, v) -plane I . When the transparency is shifted from the origin to (a, b) the Fourier transform acquires a multiplicative phase factor $\exp[-i2\pi(au + bv)]$ but this shift has no effect on the Fourier magnitude, only a phase gradient being created on the (u, v) -plane; consequently a photograph taken in the Fourier transform plane (Fig. 3a) cannot distinguish between a centred disc and a displaced one. The Hartley transform in this case responds to the shift² by acquiring a multiplicative factor $\text{cas}[2\pi(au + bv)]$, which does affect the magnitude. Figure 3b shows the Hartley intensity corresponding to the displaced disk. The jinc function factor, indicated by the concentric null rings centred on $u = 0, v = 0$, provides the origin indication mentioned above. In addition to the ring structure, this Hartley transform has fringes due to multiplication by a properly orientated cas function corrugation whose frequency and orientation are determined by the displacement. The function $\text{cas } \theta$ has a maximum at $\theta = \frac{1}{4}\pi$ and a null at $\theta = -\frac{1}{4}\pi$ and one can verify in Fig. 3b that the origin falls midway between a fringe null and a fringe maximum.

The eccentricity of the fringes illustrates the absence of symmetry that characterizes the Hartley transform and effectively enables the phase information to be encoded as amplitude. For the particular input chosen in Fig. 2, the observed Hartley transform (Fig. 3b) indicates the magnitude and direction of the displacement from the (x, y) origin and serves as an example of information that is preserved in the Hartley intensity but is, by contrast, lost in the Fourier intensity. The reciprocal nature of the transform is demonstrated by placing a properly adjusted mirror at I which causes the field $H(u, v)$ to act as the input into the transformer with the output (the reconstructed object) now appearing near T . A beam-splitter placed between T and L_1 can be used to verify the correctness of the reconstructed object.

In some circumstances holography permits the encoding of optical phase by the precise spatial location of interference fringes whose spacing is of the order of the optical wavelength. The cas-function fringes in Hartley phase encoding are in general coarser, and thus tolerant to deformation, being connected with the object structure rather than with the $\frac{1}{2}\lambda$ sec i characteristic of holography. Hartley holograms with both weak and strong isophase reference beams and with oblique reference beams have been produced.

Two other optical arrangements^{2,4}, which have not yet been implemented, one involving roof prisms and one a cube corner, offer alternative Hartley transformers that have the feature of being adaptable to incoherent illumination.

This work was supported by the Office of Naval Research and by an NSF fellowship. We thank Dr Eric Gustafson for making optical equipment available.

Static compression of H₂O-ice to 128 GPa (1.28 Mbar)

R. J. Hemley*, A. P. Jephcoat*, H. K. Mao*,
C. S. Zha*, L. W. Finger* & D. E. Cox†

* Geophysical Laboratory, Carnegie Institution of Washington,
2801 Upton Street NW, Washington, DC 20008, USA

† Brookhaven National Laboratory, Upton, New York 11973, USA

The high-pressure behaviour of H₂O is of fundamental importance in both condensed matter and planetary physics^{1,2}. The hydrogen bonding in this system gives rise to a variety of phases at low pressures and temperatures (that is, <2 GPa and <300 K), including the recently discovered high-density amorphous phases³. Structural and equation-of-state^{4,5} and spectroscopic⁶⁻⁸ studies have been carried out in the 30–50 GPa range on the dense ices (ice VII and VIII), but no data are available on the properties of solid H₂O in the megabar pressure range (>100 GPa) where a variety of stable phases, including the metallic form, have been proposed⁹. Information on the properties of H₂O at these pressures has been limited to the results of shock-wave experiments, which probe the fluid phase at high pressures and temperatures¹⁰, and to theoretical statistical electron calculations¹¹⁻¹⁵. In this study we have compressed ice in a diamond-anvil cell to 128 GPa and measured the molar volume as a function of pressure by synchrotron X-ray diffraction techniques. The diffraction data are consistent with the body-centred cubic (b.c.c.) oxygen sublattice of ice VII persisting to the highest pressures of our measurements. The measured equation of state indicates that ice is less compressible at very high pressures than is suggested by recent experiments in the 30–50 GPa range, but more compressible than statistical electron and recent pair-potential models predict.

The experiments were made possible by the use of several new applications of X-ray diffraction in the megabar pressure range with the diamond-anvil high-pressure cell. The X-ray diffraction was measured by energy-dispersive scattering techniques with a polychromatic (white) synchrotron X-ray beam^{16,17}. Synchrotron radiation, with its high intensity and low beam divergence, was crucial for these experiments. High X-ray intensity is necessary to obtain high-quality diffraction data because of the small X-ray scattering cross-section of H₂O, a low- Z material. More significant, particularly above ~20 GPa, is the fact that a small beam diameter must be used to minimize the effect of pressure gradients arising from the increasing yield strength of ice⁴. Measurements with a large X-ray beam at these pressures can result in both line-broadening effects and errors in the pressure determination. In the present experiments, the X-ray beam was collimated down to the 10–30 μm range¹⁷. Finally, to measure the pressure precisely at the sample volume probed by diffraction, the recently developed X-ray-induced ruby fluorescence method was used^{18,19}. With this technique the pressure is measured from the shift of the R_1 fluorescence line from ruby chips embedded in the sample, as in the conventional ruby method²⁰. In the X-ray-induced technique, however, the luminescence from the ruby is excited by the X-rays simultaneously with the diffraction measurement, instead of separately with a laser source. The fluorescence spectra obtained by the X-ray method are identical to those obtained in previous calibration studies over the pressure range of the present experiments¹⁸⁻²¹.

Previous structural studies of ice VII have shown that this phase has a broad stability field at room temperature, extending from 2.3 GPa to at least 50 GPa⁵. The structure is characterized by a b.c.c. arrangement of oxygen ions with some degree of disorder in the protons¹. In our experiments, doubly distilled and deionized H₂O and ruby were loaded in a high-pressure cell mounted with bevelled diamonds and designed for optical

Received 27 July; accepted 29 September 1987.

1. Hartley, R. V. L. *Proc. Inst. Radio Engrs* **30**, 144–150 (1942).
2. Bracewell, R. N. *The Hartley Transform*, 3–112 (Oxford University Press, New York, 1986).
3. Bracewell, R. N., Buneman, O., Hao, H. & Villaseñor, J. *Proc. Inst. Elect. Electron. Engrs.* **74**, 1283–1284 (1986).
4. Bracewell, R. N., Bartelt, H., Lohmann, A. W. & Streibl, N. *Appl. Opt.* **24**, 1401–1402 (1985).
5. Bracewell, R. N. *J. opt. Soc. Am.* **73**, 1832–1835 (1983).

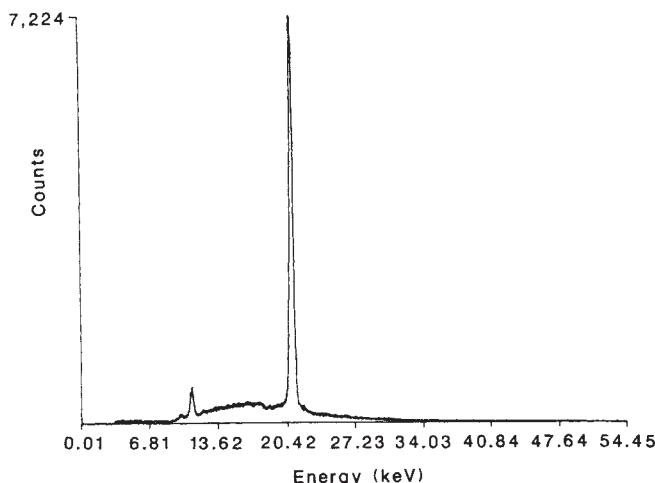


Fig. 1 Energy-dispersive X-ray diffraction pattern of H_2O -ice at 128 ± 2 GPa and 300 K. The sample, together with small chips of ruby for pressure determination, was confined in a 50- μm hole drilled into a 0.5-mm T-301 stainless steel gasket that has been preindented to 50 GPa (ref. 22). The incident X-ray beam from the synchrotron was collimated to a width of ~ 10 μm , and the diffracted X-ray beam was collected with a Si(Li) detector at a scattering angle $2\theta = 18.51 \pm 0.12^\circ$, which was calibrated with CeO_2 powder at ambient pressure. The strong line (21.07 keV) is consistent with the (110) diffraction peak of the b.c.c. oxygen sublattice. The low-energy feature at 11.15 keV is an escape peak excited in the detector and does not correspond to a diffraction line. No background (Compton scattering from the anvils) has been subtracted. The measurements were performed at beamline X13A at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory, New York.

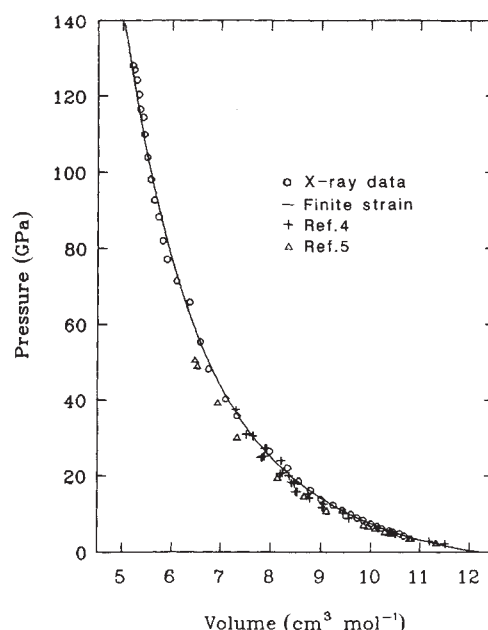


Fig. 2 Pressure-volume data and calculated room-temperature equation of state for H_2O -ice. The data of Munro *et al.*⁴ have been corrected according to the nonlinear ruby-pressure scale of Mao *et al.*²⁰ The line represents a least-squares fit of the present data to a Birch-Murnaghan finite-strain equation of state, as developed for non-quenchable high-pressure phases by Jeanloz²³. The equation of state is represented by the following parameters: the reference volume is the zero-pressure volume of ice I, $V_{01} = 19.66 \pm 0.06$ $\text{cm}^3 \text{mol}^{-1}$, and the best-fit zero-pressure volume of ice VII is $V_{02} = 12.3 \pm 0.3$ $\text{cm}^3 \text{mol}^{-1}$, with $V_{02}/V_{01} = 0.63$. The best-fit, zero-pressure bulk modulus and its pressure derivative for ice VII are $K_{02} = 23.7 \pm 0.9$ GPa and $K'_{02} = 4.15 \pm 0.07$.

and X-ray diffraction measurements at megabar pressures²². The diffraction pattern of ice VII was observed in this pressure range, consistent with previous work^{4,5}. At lower pressures ice VII has a strong propensity to form single crystals in a diamond-anvil cell, as noted in earlier investigations⁵. With increasing pressure the crystals appeared to break up to give a polycrystalline sample over the diameter of the X-ray beam, although there was still evidence for preferred orientation of the crystals. From 4.3 to 66 GPa an X-ray beam diameter of 30 μm was used. With this beam size, the uncertainty in the pressure determined by ruby fluorescence increased with pressure above 20 GPa, as a result of increasing pressure gradients across the sample. The beam size was therefore reduced to ~ 10 μm at higher pressures (>70 GPa), and care was taken to measure the pressure from a small ruby grain embedded within the diffracting volume element of the ice. This significantly improved the precision of the ruby measurement at higher pressure. In addition, with a beam size of ~ 10 μm , there was no broadening of the diffraction pattern from pressure gradients. Figure 1 shows a representative energy-dispersive diffraction pattern at the highest pressure of our measurements. The strong line can be indexed as the (110) diffraction peak of the b.c.c. oxygen sublattice of the ice VII structure. Sometimes the (220) line, which is much weaker in the measured spectra, was also observed.

Figure 2 shows pressure-volume relations calculated from the diffraction pattern assuming a b.c.c. oxygen sublattice, and Table 1 lists the numerical results. Within the resolution of the data, a continuous shift of the pattern is observed, and the results are fitted to a third-order finite-strain equation of state²³. The pressure-volume data from the earlier studies are also shown. Munro *et al.*⁴ used the earlier (linear) ruby-pressure scale, which tends to under-estimate the pressure at higher pressure (for example, 3% low at 30 GPa). We have corrected their pressures using the nonlinear ruby-pressure scale, which is currently calibrated to 180 GPa (ref. 21). The present results are in excellent agreement

with the corrected data of Munro *et al.*⁴ in the region of their maximum pressures (20–36 GPa). At lower pressures some discrepancies between the two studies are observed; these differences may be due to orientation effects resulting from the formation of large single crystals in the sample. The pressures obtained by Liu⁵ are appreciably lower at a given volume than both our results and those of Munro *et al.*⁴, particularly at higher pressures. It is possible that these differences are due to the fact that a large sample and X-ray beam were used in the previous study, and the ruby grains may have been in lower-pressure regions within the sample⁵.

Figure 3 compares the experimental equation of state determined from our data with the results of the theoretical models used in planetary studies. The Salpeter and Zapolsky¹¹ curve shows the results of a Thomas-Fermi-Dirac (TDF) calculation with electron correlation in the region of the present measurements. This calculation significantly underestimates the density in the 100 GPa range. In each of the remaining calculations^{12–15}, a TDF approach was used in the limit of very high pressures (for example, 10⁶ GPa). At low densities, experimental data (including temperature-reduced liquid-state data) were used to constrain the equation of state; various interpolation schemes between the two data sets were employed at intermediate pressures. Of these calculations, the equation of state obtained by Ree¹² agrees most closely with experiment (above 2.5 g cm^{-3}) and was specifically applied to the calculation of properties of fluid H_2O at high pressures. On the basis of TDF predictions for the low-temperature solid, Hubbard² has suggested that temperature-induced electronic excitations in fluid H_2O soften the equation of state under shock conditions. As our results are obtained at low temperatures relative to the shocked states (300 K compared with $>3,000$ K at high pressures), the extra

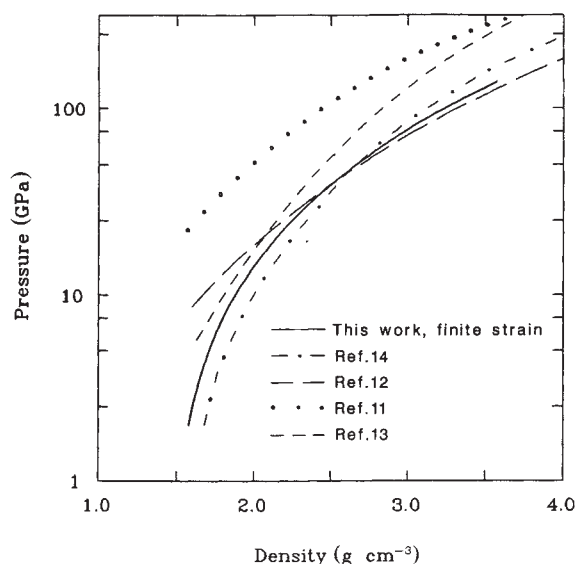


Fig. 3 Comparison of the experimental and theoretical equations of state for H_2O at high pressures and densities (see Hubbard², p. 61). The Salpeter and Zapolsky¹¹ result represents the TDF limit for dense H_2O . The calculations of Ree¹², Reynolds and Summers¹³, and Zharkov and Trubitsyn¹⁴ are semi-empirical: the curves were determined by interpolation from low-pressure experimental data and the results of quantum statistical calculations at high pressures. The Ree¹² curve is a fit to his 0 K isotherm by Hubbard and MacFarlane¹⁵. All theoretical calculations are for $T = 0$ K.

softening of the equation of state, occurring in both fluid and solid H_2O , cannot be a thermal effect but must be due to errors in the application of the TDF model. Calculations for ice VII with semi-empirical intermolecular potentials developed for the low-pressure phases of H_2O also give a significantly less compressible equation of state at high pressure (>20 GPa) than is measured experimentally²⁴.

There have been two spectroscopic studies indicating that ice VII will transform to another phase at 40–50 GPa (refs 7, 8). These investigations, made at different temperatures (100 and 300 K), suggested that the high-pressure phase is that of symmetric ice X, which is characterized by hydrogens located at the midpoint between the oxygens on the body diagonal of the oxygen b.c.c. sublattice. Hirsch and Holzapfel⁸ have observed a transition consistent with this structural change, which is first-order at 100 K and appears to involve a small volume change. Our results, obtained at 300 K, indicate that if such a transition occurs in this pressure range the transition must proceed either continuously or with a small volume change ($\leq 0.1 \text{ cm}^3 \text{ mol}^{-1}$). Some deviations in the measured pressure-volume data are apparent (for example, at 66 GPa), suggesting small discontinuities in the equation of state; however, these deviations are approximately equal to the uncertainty in the pressure. We have therefore fitted all of the data with a single equation-of-state function. It is likely that further spectroscopic measurements may provide a more sensitive probe of structural distortions of the oxygen lattice (as well as proton ordering^{6–9}) than might be apparent in this study, considering the limited number of diffraction peaks observed, and the possibility that a small splitting of the peaks may not be resolved by the energy-dispersive diffraction technique.

On the basis of these results, however, it appears that larger (reconstructive) transformations involving the oxygen sublattice do not occur in ice compressed at room temperature to 128 GPa. Polian *et al.*⁹ have suggested that ice may transform to a dense phase with the oxygens forming a face-centred cubic sublattice at ~ 100 GPa, but no such transition is observed here. In addition, pressure-induced amorphization does not occur on com-

Table 1 Pressure-volume data for H_2O -ice to 128 GPa at 300 K

P (GPa)	V_m ($\text{cm}^3 \text{ mol}^{-1}$)	P (GPa)	V_m ($\text{cm}^3 \text{ mol}^{-1}$)
4.3 ± 0.1	10.68 ± 0.03	$40.3 \pm 1.2^*$	7.11 ± 0.02
4.9 ± 0.1	10.59 ± 0.03	$48.2 \pm 1.5^\dagger$	6.75 ± 0.02
5.4 ± 0.1	10.45 ± 0.03	$55.4 \pm 2.5^\dagger$	6.58 ± 0.02
5.7 ± 0.1	10.38 ± 0.03	$65.8 \pm 2.5^\dagger$	6.36 ± 0.02
6.3 ± 0.1	10.22 ± 0.03	$71.4 \pm 0.6^\ddagger$	6.09 ± 0.02
6.7 ± 0.1	10.13 ± 0.03	77.0 ± 0.5	5.90 ± 0.02
7.5 ± 0.1	10.00 ± 0.03	82.0 ± 0.7	5.81 ± 0.02
8.4 ± 0.1	9.85 ± 0.03	88.0 ± 1.0	5.73 ± 0.02
9.0 ± 0.1	9.73 ± 0.03	92.5 ± 0.7	5.64 ± 0.01
9.9 ± 0.1	9.60 ± 0.03	98.0 ± 0.7	5.57 ± 0.01
11.0 ± 0.1	9.44 ± 0.03	104.0 ± 1.0	5.50 ± 0.01
12.3 ± 0.1	9.25 ± 0.03	110.0 ± 0.7	5.44 ± 0.01
13.8 ± 0.1	9.00 ± 0.02	113.5 ± 0.7	5.42 ± 0.01
16.2 ± 0.1	8.79 ± 0.02	116.0 ± 1.0	5.35 ± 0.01
18.6 ± 0.1	8.56 ± 0.02	120.0 ± 1.0	5.33 ± 0.01
$22.1 \pm 0.3^*$	8.34 ± 0.02	124.0 ± 1.0	5.28 ± 0.01
$26.5 \pm 0.5^*$	7.98 ± 0.02	127.0 ± 1.0	5.24 ± 0.01
$36.0 \pm 0.9^*$	7.33 ± 0.02	128.0 ± 2.0	5.20 ± 0.01

The errors listed for the pressures are those associated with the determination of the peak position of the ruby R_1 band, except where indicated. These errors arise from both peak broadening and instrumental effects. The principal contribution to the errors listed for the molar volumes is the standard error in the 2θ calibration for the energy-dispersive diffraction measurement.

* Magnitude of the quoted error is interpolated between that resulting from uncertainty in the R_1 position and that due to the pressure gradients measured above 48 GPa.

† Error is determined from the measured pressure gradient across the diameter of the X-ray beam.

‡ Smaller error follows from the reduction in beam size (see text).

pression of ice VII, an effect that occurs in the open, low-density structures (at lower temperatures)³. Finally, we note that the ice samples in the present experiments remained transparent up to 128 GPa, with no sign of band-gap closure indicative of the metallic transition^{9,25}: the ice band gap remains above the absorption edge of the diamond window of the high-pressure cell (~ 3 eV), as in solid H_2 under similar pressures²⁶. This behaviour contrasts with that of solid O_2 , in which pressure-induced electronic excitation at visible wavelengths in the 100 GPa range has been observed²⁷. From the standpoint of planetary and satellite modelling, it is clear that H_2O -ice should be considered a transparent insulating material over the pressure-temperature range of this study. Previous reports of metallic conduction in compressed ice in the 100 GPa range may be a result of poorly characterized pressures or problems associated with the electrical measurements²⁵.

The authors are grateful to the following for support of this research: the National Synchrotron Light Source, Brookhaven National Laboratory, which is supported by US Department of Energy, NASA and the Carnegie Institution of Washington. D.E.C. acknowledges support from D.O.E. Division of Materials Sciences.

Received 22 June; accepted 29 September 1987.

1. Hobbs, P. V. *Ice Physics* (Clarendon, Oxford, 1974).
2. Hubbard, W. B. *Planetary Interiors* (Van Nostrand Reinhold, New York, 1984).
3. Mishima, O., Calvert, L. D. & Whalley, E. *Nature* **310**, 393–395 (1984).
4. Munro, R. G., Block, S., Mauer, F. A. & Piermarini, G. *J. appl. Phys.* **53**, 6174–6178 (1982).
5. Liu, L.-g. *Earth planet. Sci. Lett.* **61**, 359–364 (1982).
6. Walrafen, G. E. *et al. J. chem. Phys.* **77**, 2166–2174 (1982).
7. Polian, A. & Grimsditch, M. *Phys. Rev. Lett.* **52**, 1312–1314 (1984).
8. Hirsch, K. R. & Holzapfel, W. B. *J. chem. Phys.* **84**, 2771–2775 (1986).
9. Polian, A., Besson, J. M. & Grimsditch, M. in *Solid State Physics under Pressure* (ed. Minomura, S.) 93–98 (Terra Scientific, Tokyo, 1985).
10. Mitchell, A. C. & Nellis, W. J. *J. chem. Phys.* **76**, 6273–6281 (1982).
11. Salpeter, E. E. & Zapolsky, H. S. *Phys. Rev.* **158**, 876–886 (1967).
12. Ree, F. Report UCRL-52190, Lawrence Livermore Lab., Livermore, California (1976).
13. Reynolds, R. T. & Summers, A. L. *J. geophys. Res.* **70**, 199–208 (1965).
14. Zharkov, V. N. & Trubitsyn, V. P. *Physics of Planetary Interiors*, (Pachart, Tucson, Arizona, 1978).
15. Hubbard, W. B. & MacFarlane, J. J. *J. geophys. Res.* **85**, 225–234 (1980).
16. Baublitz, M. A., Arnold, V. & Ruoff, A. L. *Rev. scient. Instrum.* **52**, 1616–1624 (1981).
17. Jephcoat, A. P. *et al. Eos* **67**, 1216 (1986).

18. Zha, C. S. *et al.* *Eos* **67**, 1216 (1986).
19. Mao, H. K. *et al.* *Bull. Am. phys. Soc.* **32**, 798 (1987).
20. Mao, H. K., Bell, P. M., Shaner, J. W. & Steinberg, D. J. *J. appl. Phys.* **49**, 3276–3283 (1978).
21. Bell, P. M., Xu, J. & Mao, H. K. in *Shock Waves in Condensed Matter* (ed. Gupta, S.) 125–130 (Plenum, New York, 1986).
22. Jephcoat, A. P., Mao, H. K. & Bell, P. M. in *Hydrothermal Experimental Techniques* (eds Ulmer G. C. & Barnes, H. L.) Ch. 19 (Wiley, New York, 1987).
23. Jeanloz, R. *Geophys. Res. Lett.* **8**, 1219–1222 (1981).
24. Grimsditch, M., Rahman, A. & Polian, A. in *Materials Res. Soc. Symp. Vol. 22* (eds Homan, C., MacCrone, R. K. & Whalley, E.) 143–146 (Elsevier, New York, 1984).
25. Liu, L.-g. & Bassett, W. A. *Elements, Oxides, Silicates - High-Pressure Phases with Implications for the Earth's Interior*, 92–94 (Oxford University Press, New York, 1986).
26. Mao, H. K., Bell, P. M. & Hemley, R. J. *Phys. Rev. Lett.* **55**, 99–102 (1985).
27. Desgreniers, S., Vohra, Y. & Ruoff, A. L. *Bull. Am. phys. Soc.* **32**, 762–763 (1987).

Internal waves and whitecaps

S. A. Thorpe, M. B. Belloul & A. J. Hall*

Department of Oceanography, The University, Southampton, SO9 5NH, UK

* Institute of Oceanographic Sciences, Deacon Laboratory, Wormley, Godalming, Surrey GU8 5UB, UK

There are numerous reports of internal waves being 'made visible' on the sea surface by their effect on the surface-wave field and the production of bands of steeper, often breaking, waves separated by zones of relatively calm water^{1–5}. The effect is sometimes quite dramatic. There are accounts of a 'low roar' as the bands of breaking waves, 'walls of white water', pass a vessel⁶. The bands are sometimes visible from aircraft⁷, on ships, radar^{8,9} and are observed from satellites^{10–15}. In the Bay of Biscay 'boils' have been reported on the sea surface in the calm zones, and appear to be related to pulses of nutrients from the thermocline¹⁶. The position of the roughest water has been commonly reported to be above the part of the internal wave in which the thermocline is most rapidly falling^{3,5} (as sketched in Fig. 1) except in very light winds ($< 3.5 \text{ m s}^{-1}$), when this appears generally to be the calmest zone¹⁶. There are, however, few simultaneous observations of surface and internal waves and none, to our knowledge, which has both quantified the frequency of the breaking waves and established their position relative to the internal waves. Here we have used an upward-pointing side-scan sonar to obtain a sub-surface view of bands of rough water and whitecaps on the sea surface associated with internal waves travelling on a pycnocline. The sonar provides a means to measure the surface currents induced by the internal waves, the position, or phase, of the surface-wave breaking relative to the internal waves, and to quantify the frequency of wave breaking. The effect of wave breaking caused by wave-current interaction is to transfer momentum from the surface waves to the currents, with a phase which tends to diminish the internal waves in the cases studied. The enhanced turbulence caused by wave breaking may be effective in mixing or eroding the pycnocline.

Explanations of these phenomena rely on the surface currents produced by internal waves. In light winds such currents concentrate surface-active films in the region above internal-wave troughs, leading to a reduction in capillary ripples and to bands of calm water¹⁷. In the more common, windy, conditions, the interactions of surface waves and the currents appear to amplify long surface waves where the current opposing them is greatest¹⁵. Shorter waves are most amplified where the horizontal gradient of the current reaches a (negative) minimum. We consider here one part of this wave-current interaction, the part which results in breaking waves and their consequent effect on internal waves and the thermocline.

Observations have been made using a two-beam side-scan sonar (248 kHz, 0.08-ms pulse length) mounted on a tripod resting on the sea bed at a depth of $\sim 35 \text{ m}$ near Oban on the west coast of Scotland¹⁸ (Fig. 1). The beams lie in two vertical planes at right angles to each other and are $\sim 1 \text{ m}$ wide at the sea surface. The overlying water is stratified by the tidal discharge of brackish water from a nearby sea-loch, Loch Etive.

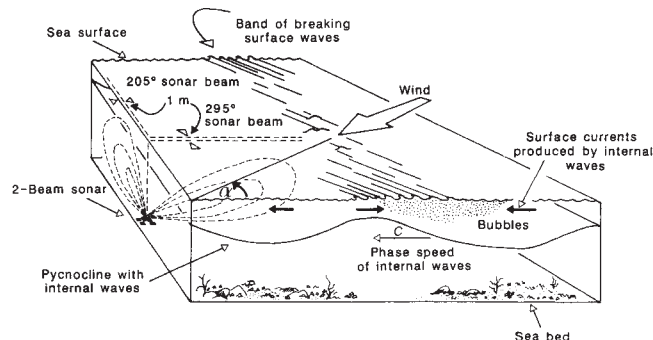


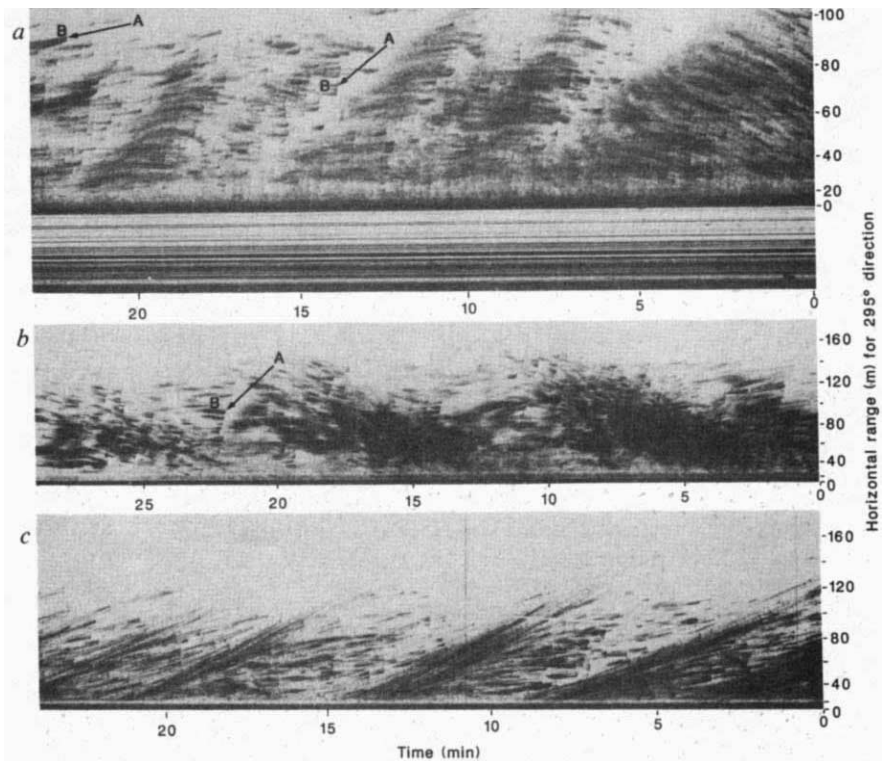
Fig. 1 Sketch showing internal waves producing near-surface currents, shown by arrows, which modulate the wind-generated surface waves to produce bands of rough water and whitecaps or breaking waves. The bottom-mounted sonar has two vertical beams at right angles.

Packets of internal waves are produced when the front heading this discharging density current is reflected from the coast $\sim 2 \text{ km}$ from the sonar position, and these are visible on the surface¹⁹. Figure 2 shows three sonograph records of these waves made apparent through the motion they produce in the sound-scattering, sub-surface clouds of bubbles created by breaking surface waves. These breaking waves usually occur in groups. Because the phase speed of the waves exceeds the group speed, successive waves approach the centre of the group, break, leave a cloud of bubbles, and then continue to advance relative to the group but with diminishing amplitude²⁰. The pattern of successive breaking waves leaves 'dashed lines' on the sonograph, indicating regions where the specular reflection from the convoluted air-water interface in the breaking waves is large (for example, at A, Fig. 2a, b), with 'plumes', that is, the bubble clouds (for example, B, Fig. 2a), being carried in the direction of the surface current. The speed, c_g , and direction of the group velocity of the surface waves are estimated from the slope of the dashed lines in two sonar beams. Using the dispersion relation, the speed provides an estimate of the wave period, $T_s = c_g/4\pi g$, where g is the acceleration due to gravity. The wave period may be determined independently from the undulating image of the surface immediately above the sonar.

The bubble plumes are advected by the surface currents produced by the wind, tide and, in these particular examples, by internal waves. The component of the current in the direction of the sonar beam is measured from the observed change in the range of the bubble clouds with time. Figure 2 shows variations in current with periods of ~ 5 – 10 min and wavelength, determined from components measured in the two sonar directions, of ~ 50 – 150 m , which are due to internal waves¹⁸. The propagation speed and direction of the waves, and the fluctuations in surface current, may all be determined from the sonograph. The fluctuations are approximately periodic about the mean current preceding the waves. This is more consistent with their approximating to a regular wave train (as sketched in Fig. 1) than to a sequence of solitary waves with unidirectional pulses of current as observed, for example, in the Sulu Sea⁷.

We have analysed the available records to determine the properties, and relative propagation directions, of the surface and internal waves and the distribution of breaking waves (see Table 1). The frequency of breaking is greatly enhanced in some regions of the internal waves, leading to the darker bands of sound reflectors (the bubble clouds), in Fig. 2. Breaking is most frequent in those regions of the internal waves where the fluctuating current tends to oppose the propagation direction of the surface waves (see also ref. 15), being typically twice as frequent there as in the zones of minimum wave breaking (the ratio r , Table 1). (The averaging over one-quarter wave periods tends to reduce the value of r , but the number of examples available

Fig. 2 Sonographs showing the motion of near-surface bubble clouds viewed by a bottom-mounted side-scan sonar. These correspond to (a) case F, (b) case B and (c) case E in Table 1 respectively. Time increases to the left in each case. *a*, Shows the complete sonograph in a beam direction of 295°. The beam is confined to a vertical plane, ~1-m wide and only targets in this beam produce significant echoes. Range is measured down the sonar beam from a point on the sea surface above the sonar. Echoes from targets near the sonar on the sea bed give rise to the constant range lines below the intense reflection from the sea surface above the sonar at zero range (this part of the record is removed in (b) and (c)). Targets which change range with time, moving in the beam direction, are due to echoes from bubble clouds (for example, at B) produced by successive breaking of surface waves mainly in groups (for example, at A). Bands of relatively dark and light areas are due to the presence, or relative absence, of bubbles resulting from the enhanced, or reduced, frequency of surface wave breaking caused by internal waves. Sonographs from only one of the two sonar beams is shown in each case. The direction of propagation of internal waves and the sonar beam directions are (b) 186° and 205°; (c) 115° and 295° respectively. Other details are given in Table 1.



and the number of breaking waves recorded make higher resolution impracticable.) It is remarkable that larger values of r , (5–6), are observed when the internal waves travel nearly at right angles to the surface waves (Table 1, examples A, D). There is a small increase in the overall frequency of wave breaking when the internal waves are present than before they arrive (*q*, Table 1). The clouds persist leaving a band of bubbles behind the region of most frequent wave breaking.

Suppose that long internal waves propagate on a pycnocline separating a surface layer of density ρ_1 , speed u'_1 , and thickness h'_1 from an underlying layer (ρ_2 , u'_2 , h'_2). If nonlinear and viscous terms, including drag on the sea bed, are neglected, the vertically integrated equations of motion in the two layers may be combined to give²¹

$$h'_1 \left(\rho_1 \frac{\partial u'_1}{\partial t} - \rho_2 \frac{\partial u'_2}{\partial t} - g(\rho_2 - \rho_1) \frac{\partial h'_2}{\partial x} \right) = \tau' \quad (1)$$

where τ' is the wind stress. We separate equation (1) into mean and fluctuating parts so that $\tau' = \tau_0 + \tau_1$, where τ_0 is the mean wind stress on the sea surface and τ_1 the fluctuation due to momentum transport in the breaking waves and u_1, u_2 are the

fluctuating parts of u'_1, u'_2 . Now using the continuity equations in the two layers to eliminate u'_2 and h'_2 we find

$$\frac{\partial^2 u_1}{\partial t^2} - c^2 \frac{\partial^2 u_1}{\partial x^2} = \frac{h_2}{h_1(\rho_1 h_2 + \rho_2 h_1)} \frac{\partial \tau_1}{\partial t} \quad (2)$$

Here h_1, h_2 are the mean values of h'_1, h'_2 , the mean currents are supposed small and to vary little on timescales corresponding to the wave period, and $c^2 = g(\rho_2 - \rho_1)h_1 h_2 / (\rho_1 h_2 + \rho_2 h_1)$. If now the fluctuating stress is related to the variable current by the phase relation $\tau_1 = (\tau/|u_1|)[u_1 \cos \phi + (\partial u/\partial x)k^{-1} \sin \phi]$, where k is the internal wavenumber and ϕ is the phase difference between the stress due to breaking waves and the current (Table 1), then a solution valid at time t such that $|\Sigma|t \ll 1$ is $u_1 = U_1 \exp(\Sigma t) \sin(kx - \sigma t)$, where the growth rate of the internal waves is $\Sigma = h_2 \tau \cos \phi / [2U_1 h_1(\rho_1 h_2 + \rho_2 h_1)]$, $\sigma = kc - \Sigma \tan \phi$, and $|\Sigma|/\sigma \ll 1$.

We may estimate τ by supposing that a fraction f of the momentum transport from wind to water is carried first to the wind waves before being transferred, as they break, into the water^{22,23}. If the factor r (Table 1), represents the modulation in the frequency of wave breaking, then

Table 1 Analysis of records to determine properties of surface and internal waves

Case	Wind speed W (m s^{-1})	α (see Fig. 1) (deg.)	Internal waves			Current magnitude U_1 (cm s^{-1})	Mean current in direction of internal waves (cm s^{-1})	Surface waves			Non-dimensional times of internal waves		
			Period, T (s)	Propagation direction (deg.)	Phase speed c (cm s^{-1})			Period T_s (s)	r	ϕ (deg.)	$\bar{N} \times 10^5$ ($\text{m}^{-2} \text{s}^{-1}$)	q	Wave 'age' T_λ/T
A	3.6	85	360 ± 83	109	18.8	8.6	~2.4	1.7	6.1	348	2.8	1.06	439
B	4.6	160	530 ± 38	185	27.2	11.1	5.5	1.9	2.8	12	0.68	1.51	-26
C	5.2	180	488	205	29.2	8.2	2.1	2.1	1.9	49	3.3	0.94	-47
D	5.4	87	532 ± 28	112	17.3	7.8	-4.0	1.3	5.0	185	2.3	No data	22
E	7.0	133	416 ± 15	158	12.9	8.5	3.2	1.9	1.7	335	5.8	1.17	-38
F*	9.0	25	444 ± 33	115	27.3	9.5	—	2.2	1.6	202	9.1	1.37	-18

The symbols denote: α , angle between directions of surface and internal wave propagation (see Fig. 1). The internal wave image on the sonograph is divided into four sections between maximum, zero and minimum currents in the wave direction. r , Ratio of the largest to the smallest number of breaking waves in these sections; ϕ , phase difference between internal wave-induced surface currents (positive in wave propagation direction) and surface breaking waves; $\bar{N}$ is the mean number of breaking waves per unit area per unit time in the internal wave region; q is the ratio of (the average number of waves breaking in a sonar beam per unit time whilst the internal waves are present) to (that measured before they arrive). All this information, except wind speed, is derived from the sonographs. In case F only one (the 295 deg.) sonar beam was in operation, and values are estimated assuming the internal waves travel down the beam direction.

$\tau = f(r-1)/[(r+1)\tau_0 \cos \alpha]$, where α is the angle between the surface and internal waves (Fig. 1) and $\tau_0 = C_D \rho_a W^2$ is the mean wind stress expressed in terms of the wind speed W , the density of air, ρ_a , and a drag coefficient, C_D . (This may be an underestimate of τ ; it ignores the changes in stress due to the observed roughening and smoothing of the sea surface.) Derived values of the non-dimensional e-folding time, $(\Sigma T)^{-1}$, where T is the internal-wave period, are given in Table 1 assuming that the air-water density ratio $\rho_a/\rho = 1.3 \times 10^{-3}$, $C_D = 1.4 \times 10^{-3}$, $\rho_2 - \rho_1 \ll \rho_1$, $f = 0.35$, (see ref. 22, for waves of average slope), $h_1 + h_2 = 35$ m and $h_1 = 4$ m (see ref. 19). In all cases where the surface waves and wind are nearly parallel to the internal waves (B, C, E, F in Table 1), breaking leads to a decay of the internal waves. When the surface waves are nearly parallel to the wave crests (A, D, Table 1) growth or decay may result, but at insignificant rates. Values of $(\Sigma T)^{-1}$ are, however, generally large, and it appears that the momentum transferred from the surface waves produced only a slow decay in the internal waves. The waves have travelled ~ 2 km from their point of generation and have propagated for a time $T_A = 2 \text{ km}/c$, given in Table 1 as a ratio of wave period T . This is less than the decay time, suggesting that little internal-wave damping will have resulted as a consequence of surface-wave breaking. (Had it been otherwise, the waves would not perhaps have been so easily detected.) Damping by this mechanism would, however, be significant in propagation distances of 10 km.

The internal-wave decay caused by surface-wave breaking may be compared with that due to the fluctuating bottom stress, $\tau_B = C_B \rho_2 u_2^2$. (C_B is the drag coefficient on the sea bed.) By analogy, this produces damping at a rate $\Sigma_B = h_1 \tau_B / [2U_1 h_1 (\rho_1 h_2 + \rho_2 h_1)]$, where U_2 is the current fluctuation amplitude in the lower layer and $h_1 U_1 = h_2 U_2$, by continuity. With typical values $U_1 = 9 \text{ cm s}^{-1}$, $r = 2$, $\phi = 180^\circ$, $\alpha = 0^\circ$, $C_B = 3 \times 10^{-3}$ (ref. 24) and f , h_1 , h_2 as before, we find $\tau/\tau_B = 0.43 \times (W)^2$ and $\Sigma/\Sigma_B = 3.3 \times (W)^2$ where W is in units of m s^{-1} . The drag produced by breaking waves is comparable to that produced by bottom friction (a factor thought to be important in dissipating internal-wave energy on the continental shelf²⁵); the damping rate may be much greater. We conclude that the effects of wave breaking may be important in damping the internal waves and, as a corollary, that internal waves travelling at right angles to the wind-wave direction may propagate further than those travelling parallel. This should be taken into account in interpreting satellite imagery of the patterns, or other measurements, of internal waves produced at the shelf break^{26,27}.

Only aspects of surface-internal-wave interaction immediately concerned with surface-wave breaking have been discussed, to the exclusion of other processes (for example refs 17, 28–30, which should be considered in a more complete analysis). Does, for example, the local enhancement of wave breaking lead to more rapid erosion of the pycnocline? Maximum wave breaking occurs at a position in the internal wave where the induced current tends to oppose the surface waves. For surface and internal waves travelling in opposite directions this will be roughly above the wave trough so that the enhanced production of surface turbulence is unlikely to have much effect on the relatively distant pycnocline. However, when the surface and internal waves travel in the same direction breaking occurs above the internal-wave crest where the pycnocline depth is reduced, and turbulence may then lead to significantly enhanced pycnocline erosion, mixing and perhaps nutrient transfer. Rates have yet to be quantified.

Received 29 July, accepted 1 October 1987.

1. Lafond, E. C. *Mar. Obs.* **37**, 10–11 (1967).
2. Fedorov, K. N. & Ginzberg, A. I. *Oceanology* **26**, 1–7 (1986).
3. Osborne, A. R. & Burch, T. L. *Science* **208**, 451–460 (1980).
4. Liu, A. K., Holbrook, J. R. & Apel, J. R. *J. phys. Oceanogr.* **15**, 1613–1625 (1985).
5. Apel, J. R., Holbrook, J. R., Liu, A. K. & Tsai, J. J. *J. phys. Oceanogr.* **15**, 1625–1651 (1985).
6. Perry, R. B. & Schimke, G. R. *J. geophys. Res.* **70**, 2319–2324 (1965).
7. Shand, J. A. *Trans. Am. geophys. Un.* **34**, 843–856 (1953).
8. Cavanie, A. *Mém. Soc. R. Sci. Liege Ser. 6* **2**, 27–44 (1972).

9. Veselov, V. M., Davydov, A. A., Shachkov, V. A., Chernyy, I. Z. & Volyak, K. I. *Izvestiya* **20**, 232–238 (1984).
10. Soules, S. D. *Deep Sea Res.* **17**, 191–195 (1970).
11. Apel, J. R., Byrne, H. M., Proni, J. R. & Charnell, R. J. *J. geophys. Res.* **80**, 865–881 (1975).
12. Trask, R. P. & Briscoe, M. B. *J. geophys. Res.* **88**, 1789–1799 (1983).
13. Alpers, W. & Salusti, E. *J. geophys. Res.* **88**, 1800–1808 (1983).
14. Hughes, B. A. & Gower, J. F. R. *J. geophys. Res.* **88**, 1809–1824 (1983).
15. Thompson, D. R. & Gasparovic, R. F. *Nature* **320**, 345–348 (1986).
16. Holligan, P. M., Pingree, R. D. & Mardell, G. T. *Nature* **314**, 348–350 (1985).
17. Lafond, E. C. in *The Sea* Vol. 1 (ed. Hill, M. N.) 731–751 (Wiley, New York, 1962).
18. Thorpe, S. A. & Hall, A. J. *Continental Shelf Res.* **1**, 353–384 (1983).
19. Thorpe, S. A., Hall, A. J. & Hunt, S. *Nature* **306**, 167–169 (1983).
20. Donelan, M., Longuet-Higgins, M. S. & Turner, J. S. *Nature* **239**, 449–451 (1974).
21. Thorpe, S. A. *J. Fluid Mech.* **63**, 509–527 (1974).
22. Mitsuyasu, H. J. *J. geophys. Res.* **90**, 3343–3345 (1985).
23. Melville, W. K. & Rapp, R. J. *Nature* **317**, 514–516 (1985).
24. Heathershaw, A. D. *Geophys. J. R. astr. Soc.* **58**, 295–430 (1979).
25. Sanford, L. P. & Grant, W. D. *J. geophys. Res.* **92**, 1828–1844 (1987).
26. Pingree, R. D., Mardell, G. T. & New, A. L. *Nature* **321**, 154–158 (1986).
27. New, A. L. *Deep Sea Res.* (in the press).
28. Gargett, A. E. & Hughes, B. A. *J. Fluid Mech.* **52**, 179–191 (1972).
29. Olbers, D. J. & Herterich, K. *Fluid Mech.* **92**, 349–379 (1979).
30. Yefimov, V. V., Kuklin, A. K. & Solov'yev, Yu. P. *Izvestiya* **20**, 560–564 (1984).

Similarity structure in the convective boundary layer of a lake

John M. Brubaker*

Centre for Water Research, University of Western Australia, Nedlands 6009 Australia and Virginia Institute of Marine Science, School of Marine Science, College of William and Mary, Gloucester Point, Virginia 23062, USA

Natural convection influences a variety of geophysical phenomena, including daily mixing cycles in the atmosphere and in surface waters, the seasonal overturn of lakes, and global climate variability. Conceptual models of convection in water are primitive and based on very limited data, particularly at small, dissipative scales. Here, I present an analysis of microscale data from the convective boundary layer of a lake, indicating a complex depth-dependence in the rate at which temperature fluctuations are dissipated. Suitably scaled, this vertical structure is remarkably similar to that found in the atmospheric convective boundary layer^{1,2}, stirred by rising 'thermals' of buoyant air heated at the Earth's surface. This comparison, consistent with recent documentation of similarity in the more uniform vertical structure of energy dissipation^{3–5}, suggests another unifying link in the structure and dynamics of convection in boundary layers of lakes, oceans and the atmosphere⁶.

The convection experiment, part of a comprehensive investigation of diurnal mixed-layer dynamics⁴, took place on a calm summer night in the central basin (horizontal and vertical scales: 2 km and 25 m) of the Wellington Reservoir, 160 km south of Perth, Western Australia. Air temperature was 12 K cooler than the water surface, barely rippled by very light winds. Surface heat flux, H , and wind stress, represented by the friction velocity u_* (Table 1), were computed⁴ from data recorded at a midlake weather station. Because the heat flux was out of the water, both the surface buoyancy flux, B , and the wind stress were sources of turbulent kinetic energy^{7,8}, with buoyancy effects dominant at depths greater than the Monin-Obukhov length $|L| = [-u_*^3/kB]$, where k is von Karman's constant. As $|L|$ was only 0.05–0.22 m (Table 1), turbulence in all but a thin surface sublayer of the 5 m mixed layer was primarily convective, or buoyancy-driven.

Eleven high-resolution temperature profiles spanning the upper mixed layer and the thermocline below were obtained in the predawn hours of 14 March 1982. Three profiles measured near 05:00 (Fig. 1) show the overall vertical thermal structure throughout the sampling period: a well-defined surface mixed layer, ~ 5 m thick, a temperature drop of ~ 0.5 K across a fairly sharp thermocline, and weak, stable stratification below the

* Correspondence should be addressed to Virginia Institute of Marine Science.

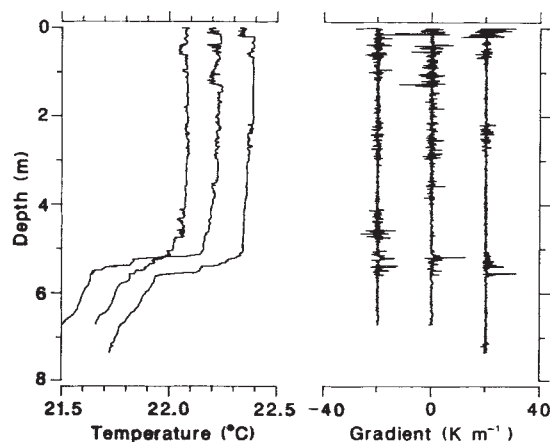


Fig. 1 Profiles of temperature and vertical temperature gradient at (left to right) 04:55, 04:59, and 05:04. Horizontal axis scales apply to the 04:59 data. Flanking temperature profiles are offset by 0.15 K, gradient profiles by 20 K m⁻¹. Positive gradients represent temperature increasing upward. The profiler design, used previously in stress-driven turbulence environments in the upper ocean¹³ and in a lake¹⁴, was implemented here in a free-rise mode with an ascent rate of 0.08 m s⁻¹ and an acquisition rate of 100 samples s⁻¹. The launch pattern was controlled so that the profiles of a set were well separated from one another and from the vessel.

thermocline. Isolated segments near the surface, colder than the ambient mixed-layer temperature, may indicate 'thermals', forming at the surface and sinking into the interior of the layer. Patches of strong microscale gradients were intermittent, but were distributed throughout the mixed layer. In the thermocline, predominantly one-sided (stable) gradients indicate little active overturning.

A key element in the dynamics of temperature fluctuations T' (considered in observational^{1,2} and modelling^{9,10} studies of atmospheric convection) is the dissipation, the rate at which thermal variance, $\langle T'^2 \rangle$, is destroyed by molecular diffusion (diffusivity D), given by $\chi = 2D\langle \nabla T'^2 \rangle$. Figure 2 shows dissipation estimates, based on all 11 gradient profiles, with a solid line connecting means of the points falling within 0.5 m bins. Large values were found near the air/water interface and near the thermocline, regions of relatively strong mean temperature gradient. In the layer interior, estimates were significantly smaller and the scatter somewhat wider. Below the thermocline, dissipation fell off rapidly with depth, to levels an order of magnitude smaller than any estimates within the convective boundary layer (CBL).

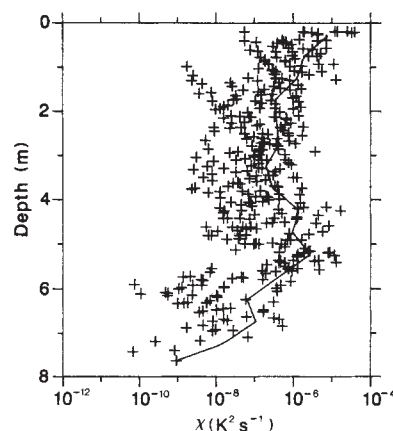


Fig. 2 Individual raw estimates of the dissipation rate of temperature variance. Each estimate corresponds to a 512-point series of temperature-gradient data, with series in each profile overlapping by 50%. All 11 profiles were used. For each series, dissipation was determined by integrating the vertical wavenumber spectrum of temperature gradients, and assuming local isotropy at dissipative scales. The integration was terminated at a cutoff wavenumber, where an abrupt drop in signal-to-noise ratio was indicated by loss of coherence between redundant amplifier channels. Solid line connects (arithmetic) mean values of estimates falling within 0.5 m bins.

In convective¹¹ (or 'mixed-layer'¹²) scaling, characteristic length, velocity, temperature, and timescales are: h (CBL thickness), $w_* = (hB)^{1/3}$, $\theta_* = H/\rho C_p w_*$, and $t_* = h/w_*$, where ρ is density, and C_p specific heat. Convective similarity for thermal dissipation may be expressed:

$$\chi/\chi_* = F(z/h) \quad (1)$$

where F is a universal, empirically determined function, and $\chi_* = 2w_*\theta_*^2/h = 2\theta_*^2/t_*$ is a natural thermal dissipation scale (the factor of two is included for compatibility with atmospheric results, where the variable $N_\theta = \frac{1}{2}\chi$ is commonly used). Throughout this study, $h \approx 5$ m and the mean value of χ_* was $1.1 \times 10^{-7} \text{ K}^2 \text{ s}^{-1}$. The raw dissipation results have been scaled by these values and pooled in bins representing depth ranges of $\frac{1}{25}h$. The resulting bin-averaged profile (Fig. 3a) has scaled magnitude of order one, well within the layer interior. The considerable increase with depth in the lower part of the CBL is in distinct contrast to the vertical distribution of energy dissipation, which has been found to be essentially uniform there in atmosphere^{1,2}, ocean^{3,5} and lake⁴ studies.

Table 1 Convective mixed layer parameters and derived scales for lake investigation and representative periods in oceanic and atmospheric experiments

u_* (m s ⁻¹)	H (W m ⁻²)	B (W kg ⁻¹)	$-L$ (m)	h (m)	w_* (m s ⁻¹)	θ_* (K)	$w_*\theta_*^2/h$ (K ² s ⁻¹)
Lake (this study)							
0.0022	216	1.2×10^{-7}	0.22	5	0.0084	0.0062	6.4×10^{-8}
0.0016	193	1.0×10^{-7}	0.10	5	0.0080	0.0058	5.3×10^{-8}
0.0013	185	1.0×10^{-7}	0.05	5	0.0079	0.0056	5.0×10^{-8}
Ocean*							
0.012†	600	2.8×10^{-7}	16	75	0.028†	0.0053†	$1.0 \times 10^{-8}†$
Atmosphere‡							
0.45		$6.9 \times 10^{-3}†$	38	1,615	2.23	0.094	$1.2 \times 10^{-5}†$

The three lines of lake data correspond to temperature sampling periods: 04:09–04:23 (4 profiles); 04:55–05:04 (3 profiles) and 06:09–06:21 (4 profiles). Friction velocity $u_* = (\tau/\rho)^{1/2}$, where τ is wind stress and ρ is water density. Surface buoyancy flux $B = g\alpha H/\rho C_p$, where g is acceleration due to gravity, α the thermal expansion coefficient, H surface heat flux (positive upward), and C_p specific heat. Layer thickness is represented by h , and the other symbols are defined in the text.

* Atlantic, 41° N 66° W, 19 January 1983, 06:00 GMT, ref. 3.

† Computed here from data given in cited reference.

‡ Northwestern Minnesota, 10 September 1973, 13:32–14:47 CDT, ref. 12.

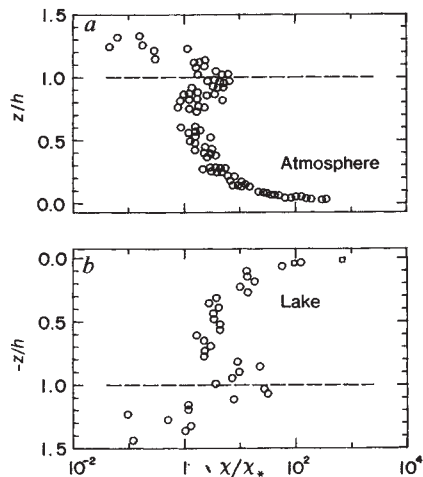


Fig. 3 Profiles of average rate of thermal variance dissipation normalized according to convective mixed-layer similarity scaling. *a*, Atmosphere profile. Circles represent bin averages of the raw estimates shown in Fig. 2 (0.2 m bins, average population 11). To extend the results nearer the surface than the 512-point series allowed, dissipation estimates based on 256-point series were determined for each microstructure record at two depths: one centred at the same depth as the shallowest 512-point series, and one at half that depth. The two squares represent the ensemble averages of these two sets of estimates. *a*, Atmosphere profile, replotted from ref. 1, based on the Minnesota and Ashchurch experiments, with dissipation determined from temperature time series at fixed levels.

This vertical structure in the lake convective layer may be compared with results from atmospheric experiments, where the scales (h and χ_* in particular) and other aspects of the setting are vastly different (Table 1). A profile of the thermal dissipation rate in the atmospheric CBL has been constructed¹, based on combined data from major experiments in Minnesota, USA¹², and Ashchurch, UK¹. This profile, scaled with convective parameters and spanning approximately the same relative vertical domain as the present observations, is replotted in Fig. 3*b*. Although there is some scatter in both the atmospheric and the lake results, the significant underlying trends in the vertical structure (appearing in mirror image) are remarkably similar and the magnitudes are of the same order.

In interpreting the extent and generality of similarity in this comparison, several points may be observed. As $|z|/h$ approaches and exceeds unity, convective similarity should not be expected in general, because turbulence structure there will be influenced by interfacial processes² and the stability of the fluid at $|z|/h > 1$, effects not represented in the scaling. Differences at small $|z|/h$ might also be expected in general, because 'surface-layer' (Monin-Obukhov) scaling^{11,12} should replace convective scaling for $|z| < |L|$, and also because $z = 0$ is a free surface for the lake, but a rigid boundary in the atmospheric experiments. Near-surface similarity observed in Fig. 3 may be attributed to the fact that $|L| \ll h$ in both settings and to the lack of substantial surface waves on the lake. With a timescale $t_* \approx 10$ min, the lake CBL would be expected to adjust rapidly enough to remain in equilibrium with its forcing, as in diurnal atmospheric convection (average t_* during sampling^{1,2,12} was 11 minutes). In a diurnally forced oceanic CBL, equilibrium convective scaling (of energy dissipation) appeared to be applicable, even with $t_* > 1$ h⁵. The stability parameter, $h/|L|$, provides further perspective. During the lake sampling, $23 < h/|L| < 100$, whereas for the majority of data at two oceanic sites⁵, $3 < h/|L| < 25$. In the cited atmospheric experiments, $h/|L|$ varied from 6 to 570 with a mean value of 144. Thus, relative to ocean and atmosphere CBL characteristics (Table 1), the lake conditions represent intermediate stability and relatively fast response time.

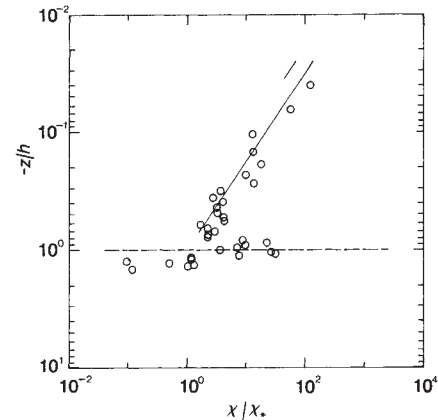


Fig. 4 log-log representation of the vertical structure of thermal dissipation, showing approximate power-law behaviour, consistent with local free convection similarity, over a portion of the mixed layer. The reference line corresponds to equation (2) with $c = 1.0$, and $c = 0.5$ is indicated by the short, parallel segment.

When $|L| \ll h$, turbulence structure over some interior sublayer of the mixed layer may be independent of h as a controlling lengthscale and independent of surface stress as well. Within such a 'local free convection'¹² or 'convective matching'¹¹ layer, functions such as F in (1) must tend to a limiting form such that h cancels out of the relation, with the result:

$$\chi/\chi_* = c|z/h|^{-4/3} \quad (2)$$

where c is a universal constant. A log-log presentation of the lake profile is shown in Fig. 4, along with a reference line corresponding to (2) with $c = 1$ (the short parallel segment represents $c = 0.5$). The decrease in dissipation rate with depth is reasonably consistent with the local free convection similarity prediction down to $|z/h| \approx 0.7$ (the limit is not well defined). Local free convection sublayers were found in atmospheric results, with estimates² for c of 0.55 based on measurements over the plains of Limagne and Beauce in France, and 0.79 for the Minnesota-Ashchurch experiments. Scaled thermal dissipation in the lake appears to be somewhat larger than has been found thus far in the atmosphere, but the difference may not be significant as it is of the same order as the difference between results found in the two pairs of atmospheric experiments.

Although factors that determine the vertical extent of similarity régimes are not well understood, the lake/atmosphere comparison presented here suggests that the observed vertical structure of thermal variance dissipation is characteristic of this class of boundary layer. The principal features: (1) peak values at the surface with rapid initial decrease with depth; (2) an internal sublayer over which the depth-dependence follows the power-law prediction of local free convection similarity; (3) scaled magnitude of order one in the mixed layer interior and (4) pronounced local maximum near the base of the mixed layer, might be expected to be found, therefore, under convective conditions in the upper ocean.

I thank J. Imberger for support and direct participation in this programme, G. Marmorino, T. Shay, and S. Thorpe for comments on an earlier version of the manuscript, and J. Kaimal for Minnesota data. The field work was supported by the Australian ARGS and AWRC.

Received 15 July; accepted 15 October 1987.

1. Caughey, S. J. & Palmer, S. G. *Q. Jl R. met. Soc.* **105**, 811-827 (1979).
2. Guillemet, B., Isaka, H. & Mascart, P. *Bound.-Layer Met.* **27**, 141-162 (1983).
3. Shay, T. J. & Gregg, M. C. *Nature* **310**, 282-285; **311**, 84 (1984).
4. Imberger, J. *Limnol. Oceanogr.* **30**(4), 737-770 (1985).
5. Shay, T. J. & Gregg, M. C. *J. phys. Oceanogr.* **16**, 1777-1798 (1986).
6. Thorpe, S. A. *Nature* **318**, 519-522 (1985).

7. Nüller, P. P. & Kraus, E. B. in *Modelling and Prediction of the Upper Layers of the Ocean* (ed. Kraus, E. B.) 143–172 (Pergamon, Oxford, 1977).
8. Fischer, H. B., List, E. J., Koh, R. C. Y., Imberger, J. & Brooks, N. H. *Mixing in Inland and Coastal Waters* (Academic, New York, 1979).
9. Zeman, O. & Lumley, J. L. *J. Atmos. Sci.* **33**, 1974–1988 (1976).
10. Isaka, H. & Guillemet, B. *Bound.-Layer Met.* **27**, 257–279 (1983).
11. Panofsky, H. A. *J. Atmos. Sci.* **35**, 272–276 (1978).
12. Kaimal, J. C. *et al. J. Atmos. Sci.* **33**, 2152–2169 (1976).
13. Dillon, T. M. & Caldwell, D. R. *Nature* **276**, 601–602 (1978).
14. Dillon, T. M., Richman, J. G., Hansen, C. G. & Pearson, M. D. *Nature* **290**, 390–392 (1981).

The dawn chorus in the great tit *Parus major* is directly related to female fertility

Ruth Mace*

Department of Zoology, Oxford University, South Parks Road, Oxford OX1 3PS, UK

Previous explanations for the dawn chorus in birds include favourable acoustic conditions¹, low foraging profitability reducing the net cost of singing^{2–4}, high risk of territorial intrusion³ and unpredictability in overnight energy requirements leading to spare reserves at dawn⁴. All these hypotheses assume that males sing either to attract mates or to defend a territory and/or that singing and foraging compete for time⁵. In the case of the dawn chorus in the great tit *Parus major* none of these assumptions apply. A peak of song at dawn only appears in late spring when territorial boundaries are established and birds are paired^{6,7}. Conflict between singing and foraging cannot explain why males with more non-foraging time in longer days at northern latitudes roost earlier than birds at southern latitudes and still sing before dawn⁸. Female emergence from her nest hole terminates the dawn chorus in the great tit⁹. I show here that changes in the duration of male dawn song are closely related to changes in female fertility.

To demonstrate a relationship between duration of dawn song and stage in the nesting cycle, observations were made on pairs of great tits nesting in Wytham Wood, Oxfordshire. I stood in a territory from civil twilight and followed the resident male for 90 minutes. Generally it was possible to follow only one male each morning although occasionally two neighbouring birds could be observed simultaneously. Data were also collected automatically on emergence times of ten females roosting in nest boxes with clocks attached.

Before relating observed changes in singing behaviour to female fertility it is necessary to know when in the laying cycle the female is fertile. Defining a precise start to the fertile period is probably unrealistic as different species store sperm for varying lengths of time¹⁰ but it is uncontroversial to state that female fertility (defined at any time as the chance of a copulation at that time fertilizing an egg) will increase as egg laying approaches. There has been some disagreement about when the fertile period ends. Although it has been suggested from behavioural evidence that it ends two or three days before the last egg is laid^{11–14}, the physiological evidence strongly favours females being potentially fertilizable until 1–2 h after laying the penultimate egg^{15,16}.

Figure 1a shows duration of dawn song over the three weeks before the onset of egg laying ($N = 14$ bird-days from five males). A multiple regression of duration of male dawn song against days from first egg, with between-bird variance removed, shows a significant increase as laying approaches. This increase in male song does not appear to be caused by any trend in the time at which females emerge from their roost holes (see legend to Fig. 1). Figure 1b shows dawn song during egg laying and the early

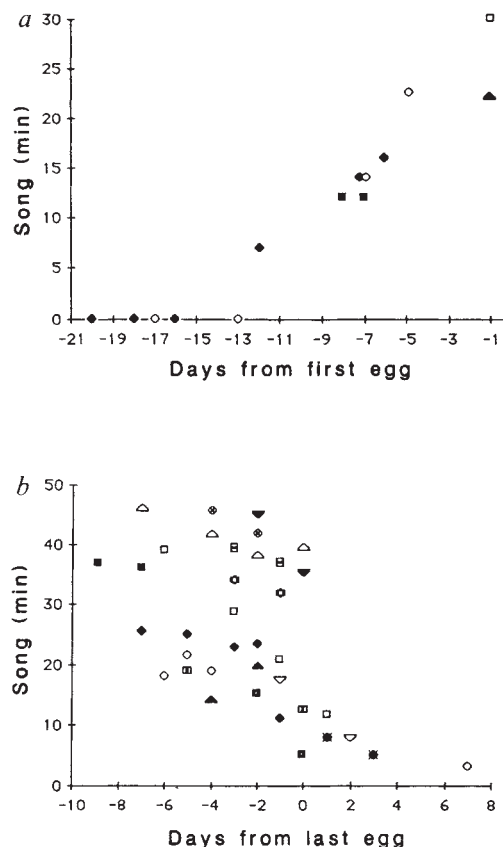


Fig. 1 Duration of male song, including pauses only of less than 15 s, in 90 min after civil twilight (sun 6° below horizon) against nesting stage. Each symbol represents a different pair. *a*, The three weeks preceding the onset of laying. Multiple regression of duration of song against days to first egg shows significant increase ($F = 45.05$, d.f. = 1, 8, $P < 0.001$; between-pair variance removed). There are insufficient days with data on both song and female-emergence time in prelaying period to remove effect of female emergence, although regression of emergence times of four females each with at least five days data in this period revealed no significant effect of days to first egg ($F = 3.00$, d.f. = 1, 30, NS; between-bird variance removed). *b*, During and shortly after laying. Multiple regression of song against days from last egg (more representative index of fertility than days from first egg as clutch sizes varied from 4 to 13) shows a significant decrease after onset of laying ($F = 27.11$, d.f. = 1, 16, $P < 0.001$; variance in female-emergence time and between-pair variance removed (female-emergence data available for only 10 birds). Similar regression with days after end of laying omitted: $F = 11.11$, d.f. = 1, 13, $P < 0.005$.

stages of incubation ($N = 37$ bird-days from 14 males). An egg is normally laid daily. A multiple regression of dawn song against days from last egg, removing variance in female emergence time and between-bird variance, gives a highly significant decrease. If data after the end of laying are omitted, a significant decline is still apparent. Hence, male dawn song increases as egg laying approaches and then decreases after the onset of laying, whereas female emergence shows no such trend.

The timing of the dawn chorus in the great tit thus suggests a direct relationship with female fertility. Characteristic behaviour at dawn during the fertile period is that the male emerges from his roost between 0 and 30 min after twilight and sings near the female's nesthole. The female emerges between 30 and 90 min after twilight, emerging ~30 min later on the mornings when she lays an egg (mean emergence times of six females were 29.3 minutes later during egg laying than prelaying, t -test: $t = 5.90$, d.f. = 10, $P < 0.001$) suggesting that eggs are laid

* Present address: Centre for Environmental Technology, Imperial College, 48 Prince's Gardens, London SW7 1LU, UK.

directly before emergence. When the female emerges, the pair dart into cover and copulate, and male song rate drops dramatically⁹. The latest copulation observed was after a female had just laid the seventh egg in a clutch of nine. A peak of copulation activity in the early morning, often with a lesser peak at dusk, has been reported for several bird species^{6,10,17-20}.

Bird song has previously been shown to have two main functions: to attract females as mates²¹ and to defend a territory²². Some species have different song types for courtship and territoriality²³ but there is no evidence for this in great tits. In this species, territorial activity is greatest several weeks before the fertile period when there is much singing but no dawn chorus⁶. Courtship, meaning a display assessed by females when deciding whom to mate with, cannot be relevant during the fertile period in these monogamous birds. But if extra-pair copulations are occurring then female assessment of males might continue after pairing causing courtship and mate guarding to be associated. My results suggest that male dawn song is a display serving to protect paternity. Increases in daytime singing²⁴ and general territoriality associated with female fertility have been observed in various species²⁵. There is growing evidence of extra-pair copulation in passerines where females had previously been assumed to be monogamous^{11-13,16,26,27}, including great tits^{6,14}. Deterring extra-pair males would be particularly important at dawn for several reasons. Physiological evidence suggests females are most fertile for 1-2 h directly after egg laying in a short 'fertilization window'^{15,28}. Furthermore low light levels, the predictable location of the female and, possibly most impor-

tant, high female motivation to mate could make successful copulations easier for sneaking males, meaning that the time when copulation normally occurs is when mate guarding would be most important, and after copulation the risk would be less.

Received 7 October; accepted 12 November 1987.

1. Henwood, K. & Fabrick, A. *Am. Nat.* **114**, 260-274 (1979).
2. Kacelnik, A. *Anim. Behav.* **27**, 237-242 (1979).
3. Kacelnik, A. & Krebs, J. R. *Behaviour* **83**, 287-309 (1982).
4. McNamara, J. M., Mace, R. H. & Houston, A. I. *Behav. Ecol. Sociobiol.* **20**, 399-405 (1987).
5. Mace, R. H. *Ardea* **75**, 123-132 (1987).
6. Hinde, R. A. *Behaviour Suppl.* **11**, 207 (1952).
7. Garson, P. J. & Hunter, M. L. *Ibis* **121**, 481-487 (1979).
8. Mace, R. H. thesis, Univ. Oxford (1987).
9. Mace, R. H. *Anim. Behav.* **34**, 621-622 (1986).
10. Birkhead, T. R., Atkin, L. & Möller, A. P. *Behaviour* **101**, 101-138 (1987).
11. Beecher, M. D. & Beecher, I. M. *Science* **205**, 1282-1285 (1979).
12. Birkhead, T. R. *Anim. Behav.* **30**, 277-283 (1982).
13. Leffelaar, D. & Robertson, R. J. *Behav. Ecol. Sociobiol.* **16**, 73-79 (1984).
14. Björklund, M. & Westman, B. *Ornis. Scand.* **17**, 99-107 (1987).
15. Sturkie, P. D. *Avian Physiology* 2nd edn (Balliere, Tindall & Cassell, London, 1965).
16. Jones, G. *Behav. Ecol. Sociobiol.* **19**, 179-185 (1986).
17. Burger, J. *Auk* **93**, 308-323 (1976).
18. Balthazart, J. *J. Zool.* **180**, 155-173 (1976).
19. Palmgren, P. *Ibis* **91**, 561-576 (1949).
20. Kok, O. B. *Condor* **73**, 348-363 (1971).
21. Eriksson, D. & Wallin, L. *Behav. Ecol. Sociobiol.* **19**, 297-300 (1986).
22. Krebs, J. R. in *Evolutionary Ecology* (eds Stonehouse, B. & Perrins, C. M.) 47-62 (Macmillan, London, 1977).
23. Catchpole, C. K. in *Acoustic Communication in Birds*, Vol. I (eds D. E. Kroodsma & E. H. Miller) 297-319 (Academic, London, 1983).
24. Greig-Smith, P. W. *Ornis. Scand.* **13**, 225-231 (1982).
25. Möller, A. P. *Oikos* **48**, 47-54 (1987).
26. Alatalo, R. V., Gustafsson, L. & Lundberg, A. *Oikos* **42**, 41-47 (1984).
27. Möller, A. P. *Behav. Ecol. Sociobiol.* **17**, 401-408 (1985).
28. Cheng, K. M., Burns, J. T. & McKinney, F. *Auk* **100**, 302-310 (1983).

Identification and distribution of 5-HT₃ receptors in rat brain using radioligand binding

G. J. Kilpatrick, B. J. Jones & M. B. Tyers

Department of Neuropharmacology, Glaxo Group Research Ltd, Ware, Hertfordshire SG12 0DJ, UK

Functional serotonin (5-hydroxytryptamine, 5-HT) receptors have been divided into three subtypes: 5-HT₁-like, 5-HT₂ and 5-HT₃ (ref. 1). Brain binding sites have been identified for both the 5-HT₁ and 5-HT₂ subtypes. Receptors of the 5-HT₃ type have been characterized on isolated peripheral tissue models such as the rat vagus nerve², guinea-pig ileum³ and isolated rabbit heart⁴. Using these models, selective 5-HT₃ receptor antagonists such as MDL 72222 (ref. 5), ICS 205-930 (ref. 6), GR38032F (ref. 7) and BRL 43694 (ref. 8) have been developed. Recently, GR38032F, MDL 72222 and ICS 205-930 have been shown to have behavioural effects in rodents and primates that undoubtedly reflect an action in the central nervous system (refs 9-11 and unpublished observations), suggesting the existence of 5-HT₃ receptors in the brain. Here we report direct evidence for the existence of 5-HT₃ receptors in rat brain tissue and their distribution, based on high affinity binding of the potent 5-HT₃ receptor antagonist ³H-GR65630 to homogenates of rat entorhinal cortex. Selective 5-HT₃ receptor antagonists and agonists inhibited binding of ³H-GR65630 with high affinities which correlated well with their actions on the rat isolated vagus nerve². Binding was differentially distributed throughout the brain with high concentrations in cortical and limbic areas.

Non-radiolabelled GR65630 (3-(5-methyl-1H-imidazol-4-yl)-1-(1-methyl-1H-indol-3-yl)-1-propanone) (Fig. 1a) is approximately 10 times more potent than GR38032 against 5-HT-induced depolarization of the rat isolated vagus nerve² (pA_{2} 9.9; Schild slope = 1.11, $n = 27$) and 2-methyl-5-HT-induced Bezold-Jarisch reflex in the cat ($ED_{50} < 0.1 \mu\text{g kg}^{-1}$ i.v.). Behaviourally,

Table 1 The affinities of various compounds that compete for ³H-GR65630 (0.2 nM) binding

	K _i (nM)		Hill number
	High	Low	
BRL 43694	0.59 ± 0.15	—	0.68 ± 0.03
Quipazine	1.4 ± 0.30	—	1.65 ± 0.25
ICS 205-930	3.1 ± 0.7	—	1.01 ± 0.01
mCPP	50 ± 5.2	—	0.97 ± 0.10
MDL 72222	55 ± 11	—	0.97 ± 0.14
Metoclopramide	360 ± 46	—	0.89 ± 0.04
Cocaine	3,600 ± 410	—	1.00 ± 0.09
GR65630	1.6 ± 0.66	32.9 ± 5.1	—
R, S-GR38032	2.9 ± 1.5	256 ± 89	—
R-GR38032	3.3 ± 0.77	220 ± 112	—
S-GR38032	5.3 ± 0.60	860 ± 19	—
5-HT	128 ± 37	—	1.47 ± 1.15
2-methyl-5-HT	87 ± 26	—	1.43 ± 0.16

Male hooded Lister rats (200-250 g) were killed and the entorhinal cortex dissected. Pooled tissue was homogenized (Ultra Turrax) in 10 vol HEPES buffer (50 mM, pH 7.4, 4 °C) and centrifuged at 4 °C and 48,000 g for 10 min. The supernatant was discarded and the process repeated for the pellet. The final pellet was suspended in 10 vol HEPES buffer. For binding, assay tubes contained 100 μl ³H-GR65630 (85 Ci mmol⁻¹, final concentration 0.2 nM, (synthesized by Dr Colin Howard, Radiochemical Group, Glaxo) in HEPES buffer, 50 μl of competing drug or its vehicle (HEPES buffer) and 100 μl of the tissue preparation (0.25-0.35 mg protein). Tubes were incubated for 30 min at 37 °C. An incubation time of 10 min was used for 5-HT and 2-methyl-5-HT because K_i values for 5-HT were markedly increased after 30 min. The incubation was terminated by rapid vacuum filtration through Whatman GF/B filters using a Brandel Cell Harvester. Filters were washed immediately with 5 × 2.5 ml HEPES buffer (room temperature). Radioactivity was assayed by liquid scintillation counting. All individual assays were carried out in replicates of three. K_i values were calculated using the computer program 'Ligand'¹³. Hill numbers are not presented for the GR compounds because the two-site analysis constrained the slopes to unity. Results are the means ± s.e.m. of at least three separate experiments.

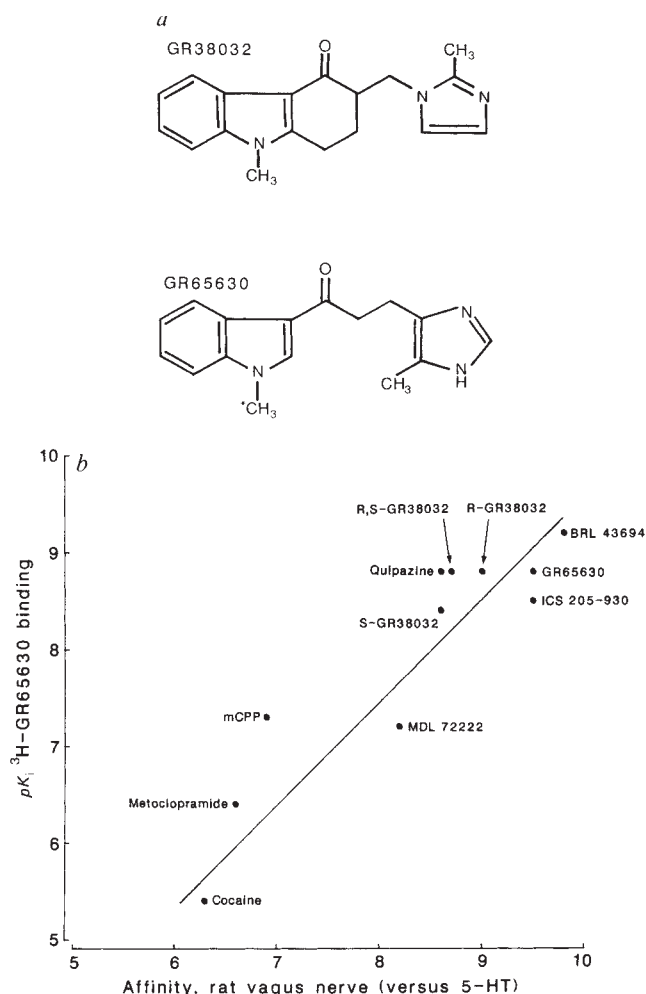


Fig. 1 *a*, The structures of GR38032 and GR65630. *, Where GR65630 was radiolabelled with tritium. *b*, Correlation of 5-HT₃ antagonist affinities to compete for $^3\text{H-GR65630}$ (0.2 nM) binding and to inhibit the 5-HT-induced depolarization of the rat isolated vagus nerve. By regression analysis the slope is 0.89 and the correlation coefficient is 0.92 ($P < 0.01$). If compounds with Hill slopes different from unity (Quipazine and BRL 43694) are excluded, the correlation coefficient is also 0.92 ($P < 0.01$). pK_i values are from Table 2. Vagus nerve experiments were performed as described by Ireland & Tyers². pA_2 values are used for the affinities of compounds which behave as competitive antagonists (cocaine metoclopramide, mCPP, R-GR38032, S-GR38032 and R,S-GR38032). pK_B values (ref. 16) are used for the other compounds.

GR65630 is active at $<10 \mu\text{g kg}^{-1}$ orally in the rat social interaction test (ref. 12, N. R. Oakley, personal communication) and inhibition of dopamine-mediated hyperactivity responses (ref. 10, R. M. Hagan, personal communication). GR65630 is inactive (at up to $1 \mu\text{M}$) against 5-HT_{1A}, 5-HT_{1B} and 5-HT₂ binding sites (unpublished observations).

In washed crude homogenates of rat entorhinal cortex, the 5-HT₃ receptor antagonist metoclopramide ($30 \mu\text{M}$) inhibited 60–70% of total $^3\text{H-GR65630}$ (0.2 nM) binding. This specific binding was linear over a wide range of tissue protein (0.1–0.8 mg per assay tube). Kinetic analysis of specific $^3\text{H-GR65630}$ binding revealed it to be a rapidly associating and dissociating ligand. Association was complete within 5 minutes ($K_{+1} = 1.0 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$). After full association, metoclopramide ($30 \mu\text{M}$) completely displaced specific binding within 2 minutes ($K_{-1} = 4.8 \times 10^{-2} \text{ s}^{-1}$). The dissociation constant (K_d) calculated from the association and dissociation rate constants (K_{-1}/K_{+1}) gave a

Table 2 Distribution of $^3\text{H-GR65630}$ (0.2 nM) binding in discrete areas of rat brain

Brain area	Specific binding (fmol per mg protein)
Cortex	
Entorhinal	11.3 ± 1.9
Retrosplenial	6.5 ± 1.1
Frontal	6.3 ± 1.0
Cingulate	6.1 ± 1.2
Temporal	5.8 ± 0.6
Occipital	4.9 ± 1.3
Parietal	4.4 ± 0.2
Lower brain regions	
Amygdala	7.2 ± 1.4
Hippocampus	4.9 ± 0.6
N. accumbens/tuberculum olf.	4.2 ± 0.5
Septum	2.1 ± 0.5
Striatum	1.3 ± 0.3
Thalamus	1.3 ± 0.5
Hypothalamus	1.3 ± 0.4
Cerebellum	0

Results are the means $\pm$ s.e.m. of three separate experiments. Experiments were performed as described in Table 1 legend. Nonspecific binding was determined by the inclusion of metoclopramide ($30 \mu\text{M}$). Brain areas are defined according to the rat brain atlas of Paxinos and Watson¹⁵.

value of 0.48 nM. Equilibrium saturation analysis (using the computer program 'Ligand'¹³) of specific $^3\text{H-GR65630}$ binding (15 concentrations, 0.05–3 nM) revealed a single saturable site of high affinity ($K_d = 0.35 \pm 0.06 \text{ nM}$, $B_{\text{max}} = 32.0 \pm 3.9 \text{ fmol per mg protein}$; mean $\pm$ s.e.m., $n = 3$). Scatchard transformation of these data was linear with a correlation coefficient of 0.98 (Fig. 2a).

When GR65630 ($10 \mu\text{M}$) was used to define nonspecific binding, two saturable sites of different affinity and capacity were evident. In these experiments 20 concentrations of $^3\text{H-GR65630}$ between 0.05 and 15 nM were used. The higher affinity site was similar to that defined by metoclopramide ($K_d = 0.36 \pm 0.14 \text{ nM}$; $B_{\text{max}} = 29.6 \pm 2.8 \text{ fmol per mg protein}$; mean $\pm$ range, $n = 2$). The lower affinity site had a K_d of $5.8 \pm 1.6 \text{ nM}$ (mean $\pm$ range, $n = 2$) and a B_{max} of $286 \pm 158 \text{ fmol per mg protein}$ (mean $\pm$ range, $n = 2$).

The selective 5-HT₃ receptor antagonists ICS 205-930, MDL 72222 and BRL 43694 as well as the non-selective antagonists quipazine, metoclopramide, mCPP and cocaine inhibited up to 70% of total $^3\text{H-GR65630}$ binding (Table 1, Fig. 2b). Hill analysis of the competition curves revealed slopes near to unity for all these antagonists except quipazine and BRL 43694. Quipazine consistently gave slopes greater than 1, while BRL 43694 slopes were consistently less than 1. GR65630 and GR38032F inhibited up to 90% of total $^3\text{H-GR65630}$ binding (Fig. 2c). Computer analysis¹³ resolved this data into two binding sites, one of high affinity and one of lower affinity (Table 1). This lower affinity site may simply be a nonspecific site recognized by GR65630 and its analogues. 5-HT₃ receptor antagonists that were structurally different did not show any affinity for this site (except MDL 72222 at high concentrations), showing that it was clearly not a 5-HT₃ receptor. The *r*- and *s*-enantiomers of GR38032 showed little selectivity, and this was also seen in functional studies (see below). The true stereoselective nature of $^3\text{H-GR65630}$ binding may only be confirmed when superior tools are available.

The affinities of the 5-HT₃ antagonists correlated well ($r = 0.92$, $P < 0.01$) with their abilities to inhibit 5-HT-induced depolarizations of the rat isolated vagus nerve² (Fig. 1b), but the correlation was not absolute. For example, mCPP and MDL 72222 were equipotent in the binding model but 20-fold

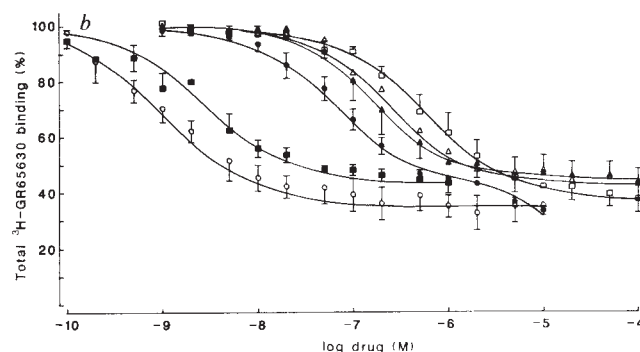
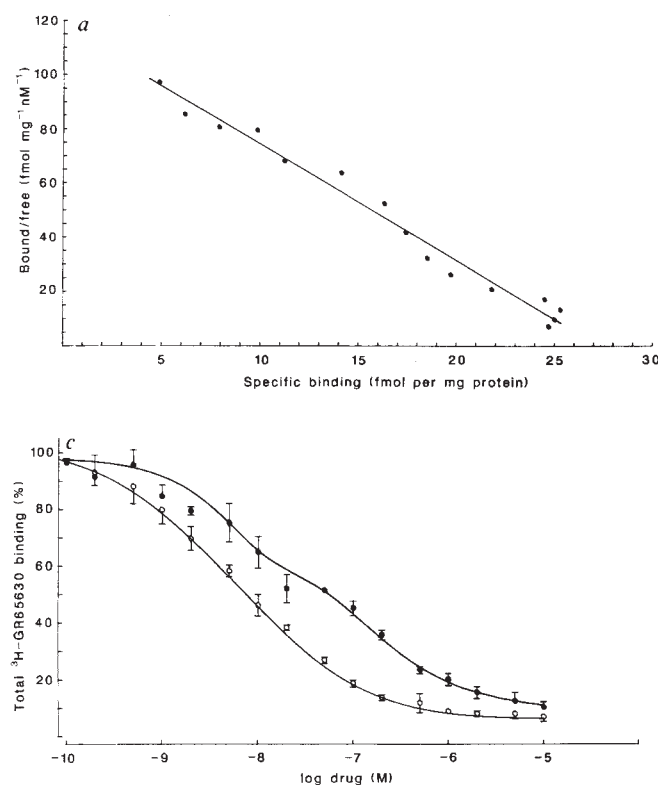


Fig. 2 *a*, Scatchard transformation of specific ³H-GR65630 (0.05–4 nM) binding to homogenates of rat entorhinal cortex. Results are from a single experiment. The K_d was 0.23 nM and the B_{max} 27.6 fmol per mg protein. Nonspecific binding was defined by metoclopramide (30 μ M). The experiment was performed as described in Table 1 legend. *b* and *c*, The abilities of 5-HT and 5-HT₃ receptor selective agonists and antagonists to compete for ³H-GR65630 (0.2 nM) binding in rat entorhinal cortex. Results are the means $\pm$ s.e.m. of three separate experiments. Experiments were performed as described in Table 1 legend. *b*, $\circ$, BRL 43694; $\blacksquare$, ICS 205-930; $\bullet$, MDL 72222; $\blacktriangle$, 2-methyl-5-HT; $\triangle$, 5-HT; $\square$, metoclopramide. *c*, $\circ$, GR65630; $\bullet$, GR38032.

different in the functional model. These anomalies may simply reflect experimental error or, alternatively, the two antagonists may have different affinities for the putative subtypes of the 5-HT₃ receptor, as proposed by Richardson and Engel¹⁴. Selective antagonists of other 5-HT receptors which were not active on the vagus nerve model, such as ketanserin, cyproheptadine, methysergide and methiothepin, did not inhibit ³H-GR65630 binding at concentrations up to 1×10^{-5} M. All ligands for various other neurotransmitter receptors which were tested against ³H-GR65630 binding were inactive (tolazoline, phen-tolamine, cimetidine, haloperidol, diazepam, noradrenaline, histamine, dopamine, atropine and GABA).

Inclusion of the selective 5-HT₃ receptor agonist 2-methyl-5-HT or 5-HT itself also inhibited up to 70% of total ³H-GR65630 binding (Table 2, Fig. 2c). In these experiments the incubation period was 10 minutes but the pK_i values for 5-HT were higher when a longer (30 min) incubation period was used. This may indicate that metabolism or oxidation is taking place. Hill analysis of the inhibition curves for 5-HT and 2-methyl-5-HT gave slope factors which were consistently greater than unity (1.2–2.0). This is atypical for agonist competition for binding to neurotransmitter receptors and may represent some form of cooperativity within a receptor complex. The fact that quipazine also showed this suggests that it may bind in a similar manner to the agonists. The addition of GTP (100 μ M) or Gpp(NH)p (100 μ M) had no effect on 5-HT, 2-methyl-5-HT or GR38032F competition for ³H-GR65630 binding. Specific ³H-GR65630 (0.2 nM) binding itself was also unaffected by the addition of

GTP (100 μ M) or Gpp(NH)p (100 μ M). This suggests that 5-HT₃ receptors may not be coupled to their effector systems through a GTP-regulated protein.

Specific binding of ³H-GR65630 was differentially localized throughout the rat brain, being highest in the entorhinal cortex (Table 2). Other areas of the cortex also showed significant binding, as did the amygdala, hippocampus and nucleus accumbens/tuberculum olfactorium. The cerebellum did not show any specific ³H-GR65630 binding. The pattern of distribution of the binding sites is striking in view of the behavioural effects in animal models of anxiety and psychosis observed with the related 5HT₃ antagonist GR38032F. The areas with the highest densities are those that are very likely to mediate these behavioural effects. For instance, effects in putative models of anxiety⁹ could involve limbic areas such as the entorhinal cortex, hippocampus and amygdala. Effects in putative models for anti-psychotic activity^{10,11} could involve dopamine-containing areas such as the nucleus accumbens/tuberculum olfactorium, amygdala and frontal cortex.

In conclusion, the specific ³H-GR65630 binding site meets the criteria for the 5-HT₃ receptor. Binding is saturable, reversible and differentially localized in the brain. Antagonists compete for the binding at affinities that correspond to their affinities in functional 5-HT₃ receptor models, whereas antagonists of other 5-HT receptors do not compete for the binding. This provides direct evidence for the existence of 5-HT₃ receptors in the brain.

We thank Dawn Seal and Lisa Taylor for technical assistance.

Received 4 September; accepted 23 October 1987.

- Bradley, P. B. *et al.* *Neuropharmacology* **25**, 563–576 (1986).
- Ireland, S. J. & Tyers, M. B. *Br. J. Pharmac.* **90**, 229–238 (1987).
- Buchheit, K. H., Engel, G., Mutschler, E. & Richardson, B. P. *Naunyn-Schmiedeberg's Arch. Pharmac.* **329**, 36–41 (1985).
- Fozard, J. R., Mobarok Ali, A. T. M. & Newgrosh, G. *Eur. J. Pharmac.* **59**, 195–210 (1979).
- Fozard, J. R. *Naunyn-Schmiedeberg's Arch. Pharmac.* **326**, 36–44 (1984).
- Richardson, B. P., Engel, G., Donatsch, P. & Stadler, P. A. *Nature* **316**, 126–131 (1985).
- Brittain, R. T. *et al.* *Br. J. Pharmac.* **90**, 87P (1987).

- Fake, C. S., King, F. D. & Sanger, G. J. *Br. J. Pharmac.* **91**, 335P (1987).
- Jones, B. J., Oakley, N. R. & Tyers, M. B. *Br. J. Pharmac.* **90**, 88P (1987).
- Hagan, R. M. *et al.* *Eur. J. Pharmac.* **138**, 303–305 (1987).
- Costall, B., Domeney, A. M., Kelly, M. E., Naylor, R. J. & Tyers, M. B. *Br. J. Pharmac.* **90**, 89P (1987).
- File, S. E. *J. Neurosci. Meth.* **2**, 219–238 (1980).
- Munson, P. J. & Rodbard, D. *Analyt. Biochem.* **107**, 220–230 (1980).
- Richardson, B. P. & Engel, G. *Trends Neurosci.* **9**, 424–428 (1986).
- Paxinos, G. & Watson, C. *The Rat Brain in Stereotaxic Coordinates* (Academic, New York, 1982).
- Rang, H. *Proc. R. Soc., Lond. B* **164**, 488–510 (1966).

Non-linear dynamics of cardiac excitation and impulse propagation

Dante R. Chialvo & Jose Jalife*

Department of Pharmacology, SUNY Health Science Center at Syracuse, Syracuse, New York 13210, USA

Extremely complex frequency-dependent patterns of excitation and impulse propagation can be shown in cardiac tissues. Such complex behaviour can be analysed using methods derived from chaos theory¹⁻⁴, which is concerned with the non-linear dynamics of deterministic systems that have irregular periodicities as well as an exquisite sensitivity to the initial conditions. We report here that the general response patterns of non-oscillatory cardiac conducting tissues, when driven rhythmically by repetitive stimuli from their surroundings, are similar to those of other deterministic systems showing chaotic dynamics. Such patterns include phase locking, period-doubling bifurcation and irregular activity. We have used electrophysiological techniques and analytical arguments to explain this unforeseen behaviour and to provide some key information about its mechanisms. The study of these dynamics is of general application to the understanding of disordered phenomena in excitable media, and may provide new insight about the origin of fatal cardiac arrhythmias.

Chaos theory focuses on the analysis of a behaviour that, while apparently random, has limited boundaries and can be quite stable. On the other hand, the term 'chaotic' has been used by cardiac electrophysiologists to describe fibrillatory cardiac

activity in which excitation occurs in a seemingly random and unpredictable fashion⁵. Recent studies⁶ of cardiac tissues have shown some forms of irregular dynamics that resemble those observed in other deterministic systems⁷⁻⁹. Indeed, phase locking and period doubling have been shown in aggregates of embryonic heart cells undergoing self-sustaining pacemaker activity¹⁰. No previous experimental attempt, however, has been made to show locking phenomena and irregular dynamics in excitable non-pacemaker cardiac tissues.

Our studies were conducted in two groups of experiments. In the first group, ten quiescent Purkinje fibres from sheep were used to determine the quantitative aspects of the response patterns to external stimuli. Preparations were driven with depolarizing current pulses applied through a suction pipette. Recordings were obtained in each experiment to construct steady-state strength-duration curves for various basic cycle lengths (BCLs) and stimulus-response ratios (SRRs). For each curve (Fig. 1a), each data point obtained at a given duration and BCL was defined as the minimal current strength at which a stable SRR was maintained. If the stimulus-response pattern repeated periodically, the SRR was termed to be 'phase-locking' or, more accurately, 'parameter-response locking'. As shown in Fig. 1a, obtained at a BCL of 700 ms, any point falling on or above the 1:1 curve would give rise to 1:1 SRR (that is, each stimulus would be followed by one action potential). At BCLs longer than those shown in Fig. 1, the only parameter-response locking observed below 1:1 was 1:0. But as shown in Fig. 1a, (BCL = 700 ms), when the BCL was decreased to the range between 1,000 and 230 ms, the 1:1-1:0 boundary became wider, and other locking patterns emerged at intermediate levels of current strength. A more complete picture of these patterns is shown in Fig. 1b. Several different parameter-response locking regions are presented. The lines are the limits of the most important isoperiodic areas (1:1, 2:1, 3:1 and 1:0) found in

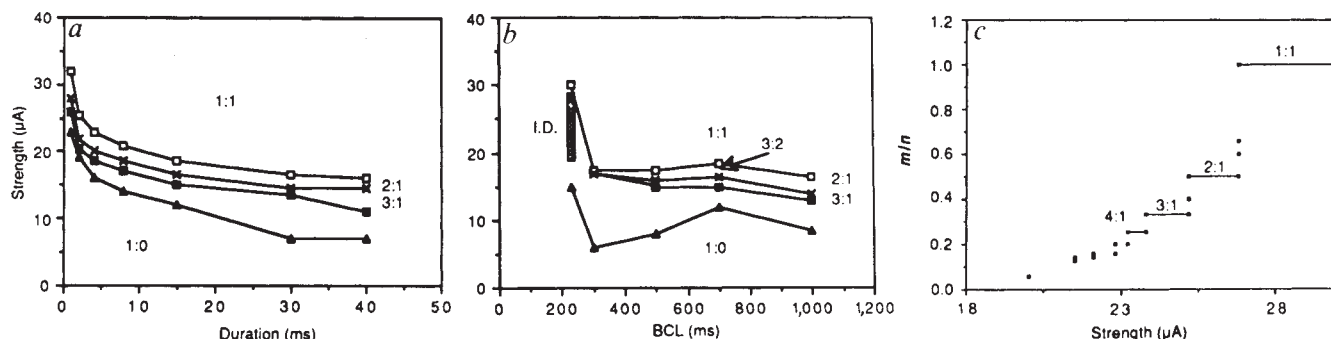


Fig. 1 a, Steady-state strength-duration curves of a Purkinje fibre from a sheep. Repetitive depolarizing-current pulses at a BCL of 700 ms. For each duration, current strength was decreased in steps of 1–2 μA beginning at levels of 1:1 locking. Symbols on each of the top three curves indicate the minimum current strength for a given pattern: open squares, 1:1; crosses, 2:1; closed squares, 3:1. Additional locking patterns between 3:1 and 1:0 (closed triangles) are not shown for simplicity. b, Parameter plane from the same experiment as in a for 15 ms-pulses at BCLs of 1,000, 700, 500, 300 and 230 ms. Curves and symbols indicate the limits of the most important isoperiodic areas (1:1, 2:1, 3:1 and 1:0). Other locking patterns occurred within very narrow regions; for instance, 3:2 locking (arrowed) between 1:1 and 2:1. The stippled area corresponds to the region of irregular dynamics. The pattern was considered to be 1:0 when, after a single action potential was initiated, at least 100 subsequent stimuli were not followed by a response. c, 'Devil's staircase'²⁴ from one Purkinje-fibre experiment. The response-to-stimulus ratio (m/n) is plotted as a function of the current strength, which was decreased in very small steps. Pulse duration and BCL were fixed at 10 and 600 ms, respectively. Beginning at 28.0 μA , 1:1 locking ($m/n = 1.0$) persisted until 26.9 μA , at which point 3:2 locking ($m/n = 0.66$) appeared. A very small ($<0.1 \mu\text{A}$) decrease in strength, however, resulted in a 5:3 pattern. At 26.8 μA , 2:1 locking ($m/n = 0.5$) was established. Further decreases in current strength yielded stable 'steps' of m/n of 0.33–0.059. If the mid-values of steps 2:1, 3:1... $n+1:1$ were connected, the resulting function would be described by the equation: $\text{strength} = (18.7 + (25.7(1/n))) - (23.1 (1/n)^2)$; $R = 1.00$.

Methods. Cardiac Purkinje fibres were dissected from the hearts of young sheep (10–25 Kg) anaesthetized with sodium pentobarbital (35 mg per Kg). Thin, externally-unbranched false tendons (outer diameter, 0.4–0.6 mm; length >10 mm) were placed in warm, oxygenated (95% O_2 , 5% CO_2) Tyrode solution of the following composition (mM): NaCl, 130; NaHCO_3 , 24; NaH_2PO_4 , 1.2; MgCl_2 , 1.0; CaCl_2 , 1.8; KCl, 4.0 and glucose, 5.6. The pH was 7.4. Group 1. Preparations were placed in a tissue bath (5 ml) and superfused with oxygenated Tyrode solution at 32 °C. At this low temperature, the ranges of the various parameter-response locking patterns were greatly amplified. Fibres were driven rhythmically by the application of constant current depolarizing pulses, delivered by a modified P6i Frederick Haer stimulator (Haer) through a suction pipette (outer diameter, 0.3 mm) at various BCLs. Current strength was measured continuously as the voltage drop across a 1 k Ω resistor in series with the suction pipette. Transmembrane potentials were recorded with a glass microelectrode filled with 3 M KCl (d.c. resistance 10–30 M Ω), connected to a high input impedance amplifier (Model 750, WP Instruments). Signals were displayed on a digital oscilloscope (Tektronix 5222) and recorded on analog tape (Model D, Vetter) for later analysis.

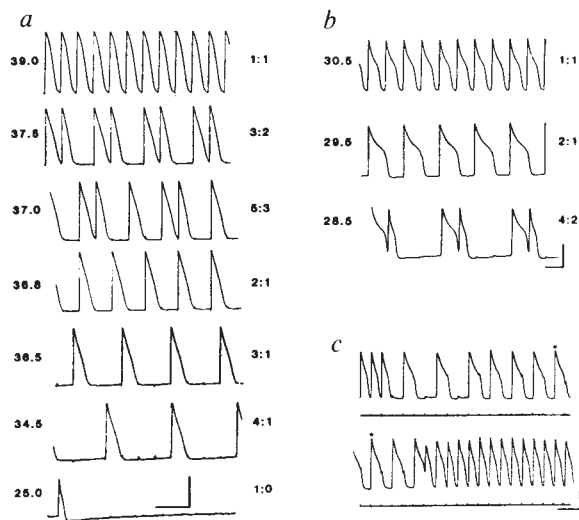


Fig. 2 Phase locking, period doubling and recurrent fluctuations in excitation of a Purkinje fibre. *a*, Transmembrane potentials obtained when current strength was decreased gradually from 39 (1:1 locking; uppermost trace) through 25 μ A (1:0 locking; bottom trace). Various other patterns (not shown) occurred between 4:1 and 1:0. After the last action potential at 25 μ A, no activity was manifest even during more than the 12 s of continuous recording shown here. Numbers at left indicate current strength in μ A; numbers at right denote parameter: response ratios. Pulse duration, 5 ms; BCL, 300 ms. Vertical calibration is 60 mV for all traces; horizontal calibration is 600 ms for all, except bottom trace (1,500 ms). *b*, Analog recordings from a different experiment (pulse duration, 15 ms; BCL, 600 ms). At 30.5 μ A, 1:1 was present. This was followed by 2:1 at 29.5 μ A. Further decreasing the current strength to 28.5 μ A, however, led directly to a 4:2 locking pattern, and not to the expected 3:1. Calibrations are: vertical, 60 mV; horizontal, 60 ms. Numbers on right and left of traces are same as in *a*. *c*, Recurrent fluctuations in the activity in the same fibre as in *b* (pulse duration, 20 ms; strength, 41 μ A; BCL, 500 ms). A stimulus monitor is shown below each trace. The last action potential (diamond) on top panel of *c* is reproduced at the beginning of bottom panel. At constant pulse magnitude and BCL, 1:1 locking is present during the first three beats. But the fourth stimulus fails to initiate a response, which leads to various unstable patterns until 1:1 is reestablished. This behaviour repeated periodically after every 40 to 50 beats. Note large prolongation of action potential duration during fourth response. Calibrations: vertical 50 mV; horizontal 1,000 ms.

this representative preparation. The largest area between 1:1 and 1:0 was 2:1, between 1:0 and 2:1 was 3:1, between 3:1 and 1:0 was 4:1, and so on. Several forms of Wenckebach-like patterns¹¹ were also manifest. Curiously, these patterns were found in areas that were proportionately narrower than those between 1:0 and 2:1. Consequently the most commonly observed stable locking of this kind was 3:2 (arrowed in Fig. 1*b*), whereas others (for example 4:3 and 5:4) were seen as unstable patterns between 1:1 and 3:2 or 2:1.

Certain types of parameter-response locking occurred within relatively narrow areas of stimulus strength and BCL. For example, 5:3 locking could be manifest as an alternation between 2:1 and 3:2 (see Fig. 2*a*). In fact, the overall hierarchy follows the so-called Farey sequence that is universally observed in oscillatory systems^{12,13}. Hence, within a given range of BCLs, the largest area between $n:m$ and $N:M$ was always $n+N:m+M$; where n and N are the number of pulses and m and M the number of responses. Thus, our data show that this structure is not restricted to self-sustaining oscillators, which was predicted by previous theoretical and experimental studies¹⁴⁻¹⁹.

Interestingly, the periodic patterns we have observed in all our experiments adhere also to another well-known universal

rule²⁰⁻²²—if in a given sequence of stimulation we define successful activation as 1 and failure as 0, the periodic pattern 1,1,0,1,0 (5:3 locking) will be found only between areas of 1,1,0 (3:2 locking) and 1,0 (2:1 locking), which is true also for longer sequences, depending on the amount of experimental noise acting to destroy such sequences. These observations suggest that, when the pulse parameters are changed, it is possible to predict the locking period (for example 2:1, 3:1, and so on following Farey's series) as well as the order of action potentials and dropped beats within a given sequence.

The experimental data presented in Fig. 1*c* show the scaling properties of the behavioural structure. In this case, the action potential (m) to stimulus (n) ratio is plotted as a function of the pulse strength. Clearly, the plot resembles the 'devil's staircase' of phase locking in oscillatory systems²³. In a strict sense, however, if such a scaling behaviour is predictable as in other deterministic systems²⁴, a function joining the mid-values of the steps 2:1, 3:1... $n+1:1$ should be close to a power law of 2. Analysis of our results confirmed the existence of that function.

The hierarchical structure described above was lost at values of BCL and strength such as those within the stippled area of Fig. 1*b*. In this region, we found irregular dynamics (ID) coexisting with parameter-response locking. Normally, when 1:1 locking was lost as pulse strength was decreased in small steps, the sequence of stable phase-locking patterns was 3:2, 5:3, 2:1, 3:1, 4:1, ... 1:0 (Fig. 2*a*). But in some cases, lowering the stimulus strength changed the pattern from 1:1 directly into 3:1 or 4:1 rather than the expected 2:1. In other cases, such as that illustrated in Fig. 2*b*, decreasing the current-pulse strength in very small steps from the 1:1 domain yielded the sequence 1:1 $\rightarrow$ 2:1 $\rightarrow$ 4:2. This can be regarded as a period-doubling bifurcation as has been described in other systems¹. Other more typical examples of period-doubling bifurcation are also present, including a 2:2 pattern in which a stimulus is always followed by an action potential whose amplitude, duration or latency (see Fig. 3*a*) alternates on a beat-to-beat basis.

In yet other cases, we found recurrent fluctuations in the behaviour. Indeed, as shown in Fig. 2*c*, under some accurately controlled conditions of fixed stimulus magnitude and BCL, 1:1 locking was lost and the sequence became: 1:1... 2:1 $\rightarrow$ 3:1, 3:1 $\rightarrow$ 2:1... 2:1 $\rightarrow$ 1:1... 2:1 $\rightarrow$ 3:1 and so on. Three different frequencies coexist during this activity: the frequency of stimulation, the frequency of the fibre response and the frequency of repetition of the overall behaviour. In our experiments, this particular type of activity may have depended on at least two dynamical factors. First, the fibres must have had supernormal excitability at very early intervals²⁵, and second, the prolongation of the action potential duration after the first dropped beat must have occurred rapidly. Similar dynamics were found also near other stable phase-locking regions (for instance, 2:1 and 3:1), suggesting that recurrent fluctuation 'surrounds' the stable parameter-response locking regions.

The strength versus BCL plot of Fig. 1*b* can be considered as the parameter plane of our non-linear system. In such a plane, the irregular dynamics are more often observed within narrow regions at the upper-left margin, rather than at the lower phase-locking regions. Parameter-space analysis has shown that non-linear driven oscillators also show this proportionality^{26,27}.

In a second group of eight experiments, the dynamic behaviour of our preparations was studied as a spatial phenomenon during action-potential propagation along a linear non-homogeneous cell strand. In this case, the parameters that were used to induce the various dynamics were changes in the BCL and in the effective conduction velocity along the strand. Unbranched, quiescent Purkinje fibres from sheep were placed in a three-chambered tissue bath²⁸ and driven by surface stimuli applied at one end of the fibre. Effective conduction velocity was measured between two microelectrodes, one proximal (P) and the other distal (D) to the stimulating electrode. When all three segments were perfused with normal Tyrode, impulses

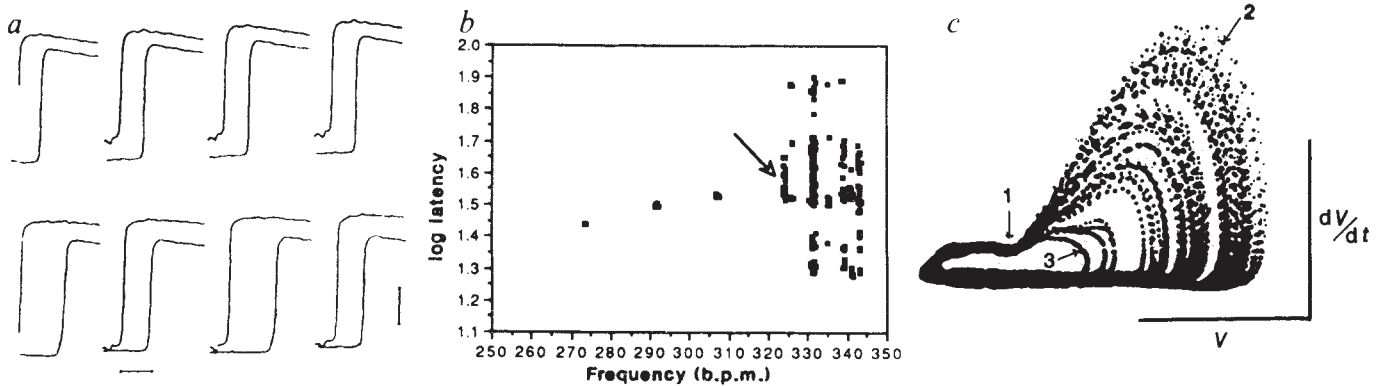


Fig. 3 *a*, Period doubling of cardiac impulse propagation. In both panels, the top trace is proximal (P), and bottom trace is distal (D) action potential recording from a linear Purkinje fibre placed in a three-chamber bath²⁸. The top panel shows frames (duration 6 ms) of four consecutive beats at a BCL of 212 ms. P-D latency was consistent. A similar sequence is shown in the bottom panel at a BCL of 208 ms. In this case, P-D latency alternated between 1.6 and 2.9 ms. P-D distance, 6.2 mm. Conduction velocity at BCL, 1,000 ms was relatively slow (4.4 m s^{-1}) because of slight compression on the fibre by the two rubber membranes bounding the central chamber (1.5 mm). Calibrations: vertical, 40 mV; horizontal, 2 ms. *b*, Irregular dynamics following period doubling of P-D latency in a three-chamber bath experiment. Abscissa, P frequency; ordinate, log of P-D latency. P-D latency increased monotonically as frequency was increased. Latency doubling began to occur at the critical value of 322 b.p.m. (arrow), with the 'arms' of the bifurcation separating gradually as the frequency was further increased. Irregular dynamics developed between 325 and 345 b.p.m. During more than 100 beats in which irregular dynamics were recorded, the sequence of latencies was always long-short-long, and so on. Note that certain intermediate values of P-D latency were never recorded. *c*, Phase plane trajectories (V = membrane potential; dV/dt = first derivative of membrane potential) of the D cell, taken from the same experiment as in *b* during irregular dynamics at 332 b.p.m. Inter-electrode distance, 8.7 mm; heptanol (1.5 mM) superfusion in central chamber (1.48 mm in length); P-D latency at BCL = 1,000, 4.0 ms (conduction velocity, 2 m sec^{-1}). Calibrations: vertical, 200 V s^{-1} ; horizontal, 40 mV.

Methods. Group 2. A three-compartment bath²⁸ was used for these experiments. The fibres were threaded through the holes of two latex membranes that bounded the central chamber of the bath, and were superfused with Tyrode solution at 37°C . Preparations were driven through a pair of surface bipolar electrodes on one end of the fibre. Transmembrane potentials were recorded simultaneously with two microelectrodes, each located in one of the outer chambers. All signals were amplified, monitored and stored as described for Group 1 (see Fig. 1 legend). For all experiments, data were analysed by playing back the tape-recorded signals on the digital oscilloscope through a low-pass filter, and sending them to an analog X-Y plotter (Omniographic, Houston Instruments). Latency versus frequency plots (see Fig. 3*b*) were constructed also directly from the tape-recorded signals. P-P intervals and P-D latencies were measured with an input-output and A-D conversion module (models 64IF22 and 64IF/ADC0816, Schnedler Systems, Inc.) interfaced with a microcomputer (Commodore C128). Zero voltage crossing of the P or D signal started and stopped the computer clock (precision $10 \mu\text{s}$). Phase-plane trajectories were obtained off-line by feeding the recorded voltages into an analog differentiator (linear between 1 and 500 V s^{-1}), and plotting the resulting signal against voltage on the digital oscilloscope. Variable external sampling rate (faster during the upstrokes, slower during repolarization) was used to save oscilloscope memory (limit, 1,024 samples), thus allowing the superposition of about 100 trajectories on each frame.

initiated at a frequency of 1 per s propagated in a one-to-one manner, with conduction velocities greater than 5 m s^{-1} . Stimulation at increasing rates led to a progressive increase in P-D conduction time (latency), until the pattern evolved directly and predictably into 2:1. When the central segment (1–2 mm long) was slightly compressed or superfused with 1.25 to 1.5 mM of the electrical uncoupler heptanol²⁹, 1:1 conduction at a rate of 1 per s became slower than in the control. At intermediate levels of conduction impairment (with a conduction velocity of 5 to 1.5 m s^{-1}), stimulation at increasing rates also led to a progressive increase in P-D latency, until 2:1 locking occurred. Under these conditions, however, the 2:1 pattern was preceded by period doubling. Figure 3*a* shows an example. Oscilloscope tracings of four consecutive P and D action potentials are shown in each panel during the initial 6 ms after the corresponding stimulus. In the upper panel, at a BCL of 212 ms, 1:1 locking was observed with a constant P-D latency of 1.7 ms. As the BCL was reduced in 1 ms steps, the P-D latency began to alternate on a beat-to-beat basis, with a gradual increase in the difference between brief and long latencies. In the lower panel, at a BCL of 208 ms, P-D latency alternated between 1.6 and 2.8 ms.

Figure 3*b* shows the results of a more complete experiment in which the first period-doubling bifurcation (2:2) was followed by more irregular dynamics, before 2:1 locking occurred. The P-D latency (ordinate) is plotted as a function of the driving frequency (abscissa). Clearly, as the driving rate was increased, P-D latency also increased until, at 322 beats per min (b.p.m.), a bifurcation was reached (arrow) at which long and brief

latencies began to alternate. From this point on, all increments in pulse frequency were made in 1 ms^{-1} steps and at least 100 beats were plotted for each step. The period was 2 between the initial bifurcation and 325 b.p.m., where latency again became unstable. Higher periodicities began to emerge as the frequency changed from 330 to 345 b.p.m. Even though, under these conditions, the variation in latency was large and complex, it was not random, as certain values of latency were never visited. This suggests a chaotic structure³⁰.

The results in Fig. 3*c* are presented in the form of phase-plane trajectories. One hundred responses are plotted at a driving rate of 332 b.p.m. Two important dynamical features are shown by this experiment. The first one, already alluded to in the latency domain, is that some orbits are never visited, as is clear by the structure of the state-space portrait (Fig. 3*c*). The second one is a didactic example of the property known as 'sensitive dependence on the initial conditions'. During the n th beat, the state variables (membrane voltage, V , and dV/dt) are in the sub-threshold excursion represented by point 1. Ten milliseconds later, the state variables have moved to point 2. In the next beat ($n+1$), the variables are again very close to, but not precisely at point 1. As a consequence of this small difference in the initial conditions, there is a strong divergence in the subsequent trajectory which, after 10 ms, leads the state variables to point 3. This behaviour is very similar to that described for other deterministic systems that show chaotic dynamics³¹.

Finally, in these experiments, when the degree of conduction impairment induced by segmental compression or by superfusion with heptanol was high (conduction velocity $< 1.5 \text{ m s}^{-1}$

at BCL = 1,000 ms), various sequences of parameter-response locking were found as the driving frequency was changed. These sequences were very similar to those described for the first group of experiments, and also followed Farey's universal progression of phase-locking patterns.

In conclusion, parameter-response locking, irregular dynamics and chaotic-like behaviour can occur in cardiac Purkinje fibres that do not undergo spontaneous activity. This may have important implications from the theoretical as well as from the experimental and clinical points of view. Theoretically, the surprising behaviour of our preparations raises two open questions directed to mathematicians and theoretical physicists. Is self-sustaining oscillation fundamental to the demonstration of chaotic dynamics? If not, can the chaotic dynamics of a non-oscillatory system that follow universal rules be described by the same theoretical techniques^{32,33} as those used in the analysis of driven non-linear oscillators? For the experimental and clinical cardiologist, the implication of such dynamics for the study of cellular mechanisms of cardiac-rhythm disturbances and for the development of sudden cardiac death³⁴ is extremely interesting and deserves to be explored. Nonlinear dynamics has helped to suggest new experimental approaches in the analysis of normal electrophysiology of the heart and its alterations³⁵. By focusing on these properties it might be possible to reach an understanding of the mechanisms of a number of rate-dependent conduction disturbances^{36,37}. By utilizing biological models of cardiac electrical activity to study the transition from regular to irregular (chaotic?) dynamics, new insight might be gained into the origin of erratic, sometimes fatal arrhythmias observed in the clinic.

This work was supported by the NIH. We thank D. C. Michaels, M. Delmar and J. Anumonwo for helpful discussions, P. Chiale for his help in some of the experiments, and O. Piro for sharing his unpublished data with us.

Received 28 August; accepted 29 October 1987.

- Holden, A. V. *Chaos* (Princeton University Press, 1986).
- Cvitanovic, P. *Universality in Chaos* (Hilges, Bristol, 1984).
- Hao, B.-L. *Chaos* (World Scientific, Singapore, 1984).
- Devaney, R. L. *An Introduction to Chaotic Dynamical Systems* (Benjamin Cummings, Menlo Park, 1986).
- Moe, G. K., Rheinboldt, W. C. & Abildskov, J. A. *Am. Heart J.* **67**, 200-220 (1964).
- Guevara, M. R., Glass, L. & Shrier, A. *Science* **214**, 1350-1353 (1981).
- Simoyi, R. H., Wolf, A. & Swinney, H. L. *Phys. Rev. Lett.* **49**, 245-248 (1982).
- Gollub, J. P. & Benson, S. V. *J. Fluid Mech.* **100**, 449-470 (1980).
- Testa, J., Perez, J. & Jeffries, C. *Phys. Rev. Lett.* **48**, 714-717 (1982).
- Glass, L., Shrier, A. & Belair, J. in *Chaos* (ed. Holden, A.) 237-256 (Princeton University Press, 1986).
- Zipes, D. P. *Circulation* **60**, 465-472 (1979).
- Arnold, V. I. *Geometrical Methods in the Theory of Ordinary Differential Equations. Section II* (Springer, New York, 1983).
- Glass, L. & Perez, R. *Phys. Rev. Lett.* **48**, 1772-1775 (1982).
- Guevara, M. R., Ward, G., Shrier, A. & Glass, L. *Comp. Cardiol.* **11**, 167-170 (1984).
- Feingold, M., Gonzalez, D., Piro, O. & Vitturo, H. *Phys. Rev. Lett.* (submitted).
- Harmon, L. P. *Kybernetik* **1**, 89-101 (1961).
- Guttman, R., Feldman, L. & Jakobsson, E. *J. Membrane Biol.* **56**, 9-18 (1980).
- Glass, L., Guevara, M. R. & Shrier, A. *Ann. N.Y. Acad. Sci.* **584**, 168-178 (1987).
- Matsumoto, G. *et al. Phys. Lett. A* **123**, 162-166 (1987).
- Sato, S. *Kybernetik* **11**, 208-216 (1972).
- Nagumo, J. & Sato, S. *Kybernetik* **10**, 155-164 (1972).
- Yoshizawa, S., Osada, H. & Nagumo, J. *Biol. Cybern.* **45**, 23-33 (1982).
- Mandelbrot, B. B. *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982).
- Kaneko, K. *Kollege of Tori and Genesis of Chaos in Dissipative Systems* (World Scientific, Singapore, 1986).
- Spear, J. F. & Moore, E. N. *Circulation Res.* **35**, 782-792 (1974).
- Gonzalez, D. & Piro, O. *Phys. Rev. Lett.* **50**, 870-872 (1983).
- Guevara, M. R. & Glass, L. *J. Math. Biol.* **14**, 1-23 (1982).
- Jalife, J. & Moe, G. K. *Circulation Res.* **39**, 801-808 (1976).
- Johnston, M. F., Simon, S. A. & Ramon, F. *Nature* **286**, 498-500 (1980).
- Rapp, P. E. *et al. in Lecture Notes in Biomathematics* **66**, *Nonlinear Oscillations in Biology and Chemistry* 175-205 (1986).
- Vidal, C. in *Nonlinear Phenomena in Chemical Dynamics* (eds Vidal, C. & Pacault, A.) 49-62 (Springer, Berlin, 1981).
- Guckenheimer, J. *IEEE Trans. Circuits and Systems* **30**, 586-591 (1983).
- Borenblatt, G. I., Ioos, G. & Joseph, D. *Nonlinear Dynamics and Turbulence* (Pitman, London, 1983).
- Ritzenberg, A., Adam, D. & Cohen, R. *Nature* **307**, 159-161 (1984).
- Winfree, A. T. *When Time Breaks Down. The Three-Dimensional Dynamics of Electrochemical Waves and Cardiac Arrhythmias* (Princeton University Press, 1987).
- Rosenbaum, M. B., Elizari, M. V., Levi, R. J. & Nau, G. J. *Chest* **63**, 678-688 (1973).
- Halpern, M. S., Nau, G. J., Chiale, P. A., Elizari, M. V. & Rosenbaum, M. B. in *Frontiers of Cardiac Electrophysiology* (eds Rosenbaum, M. B. & Elizari, M. V.) 465-487 (Nijhoff, The Hague, 1983).

Phosphorylation fails to activate chloride channels from cystic fibrosis airway cells

Robert A. Schoumacher*, Richard L. Shoemaker†, Dan R. Halm†, E. Ann Tallant‡, Robert W. Wallace‡ & Raymond A. Frizzell†

* Departments of Pediatrics, † Physiology and Biophysics, and ‡ Pharmacology and Gregory Fleming James Cystic Fibrosis Research Center, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA

Chloride impermeability of epithelial cells can account for many of the experimental and clinical manifestations of cystic fibrosis (CF)^{1,2}. Activation of apical-membrane Cl⁻ channels by cyclic AMP-mediated stimuli is defective in CF airway epithelial cells^{3,4}, despite normal agonist-induced increases in cellular cAMP levels^{4,5}. This defect in Cl⁻ channel regulation has been localized to the apical membrane by exposing the cytoplasmic surface of excised membrane patches to the catalytic subunit (C subunit) of cAMP-dependent protein kinase and ATP. In membranes from normal cells, C-subunit activated Cl⁻ channels with properties identical to those stimulated by cAMP-dependent agonists during cell-attached recording. Activation by the C subunit was not observed in CF membranes, but the presence of Cl⁻ channels was verified by voltage-induced activation. The failure of the C subunit to activate the Cl⁻ channels of CF membranes indicates that the

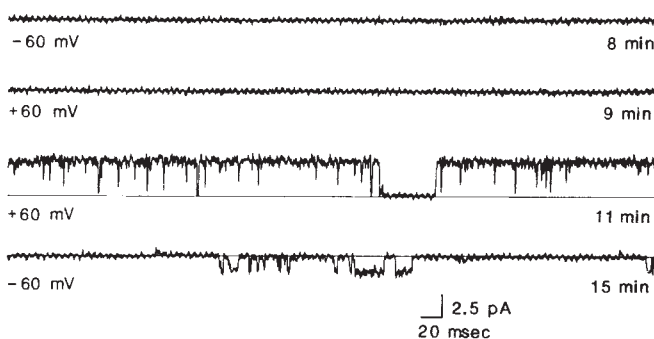
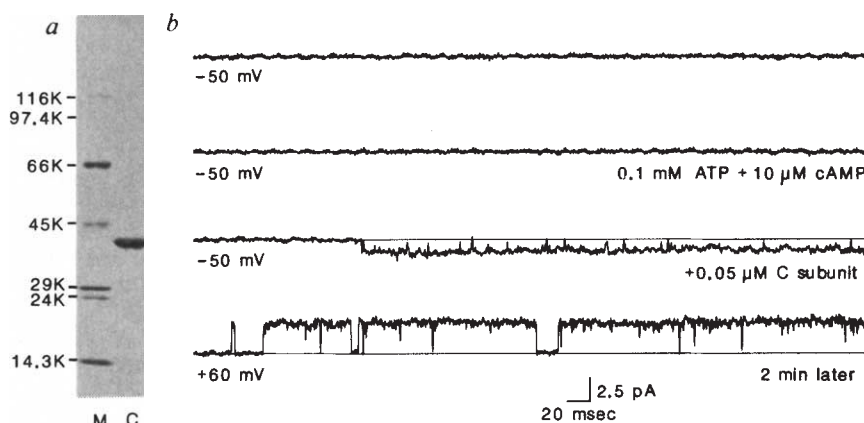


Fig. 1 Time course of voltage-induced Cl⁻ channel activation. Successive records were obtained at times shown at the right margin and at holding potentials given at the left margin. Membrane voltage is expressed with respect to pipette as reference; upward current deflections are outward currents, carried by Cl⁻ flow from the outside to cytoplasmic side. Top trace: no activity was observed for 8 min after excision; membrane voltage was -60 mV; Trace 2: no activity at 9 min, 1 min after voltage was switched to +60 mV; Trace 3: outward Cl⁻ current at 11 min, +60 mV; Trace 4: inward Cl⁻ current at 15 min, -60 mV.

Methods. Single-channel recordings were obtained from membrane patches excised from normal and CF human tracheal epithelial cells in primary culture, using methods described previously³. Tissues from seven normal and three CF patients were studied. Recordings were obtained from the upward-facing membranes of single cells or confluent islands of 5-50 cells. The results were similar, suggesting that cultured airway cells attached to substrate express the differentiated properties of the apical membrane that are of interest in this study. The pipette solution was (mM): 160 NaCl, 2 CaCl₂, 5 HEPES, pH 7.2; the bath contained: 150 NaCl, 5 MgCl₂, 0.1 CaCl₂, 1.0 NaEGTA, 5 HEPES, pH 7.2; calculated free Ca ions ~20 nM. Current records were obtained at a filter bandwidth of 1 kHz. Open channel probability (P_o) was calculated as the fractional area under open channel records, with the assumption that the maximum number of observable unitary current levels at any voltage represents the number of channels in the patch.

Fig. 2 Chloride-channel activation by the C subunit in a normal membrane patch. *a*, SDS-PAGE of the C-subunit preparation used in these studies. Lane M shows the relative molecular mass (M_r) markers; lane C shows the C-subunit preparation, with a single band at 41K. *b*, Time course of activation. Top trace: the excised patch was electrically quiet at -50 mV for ~ 2 min; Trace 2: ATP plus cAMP were added to the bath; no activity was observed for another 8 min at -50 mV; Trace 3: inward currents were evoked by the C subunit 3 min after its addition to the bath (ATP and cAMP still present); Trace 4: outward Cl^- currents at $+60$ mV following C-subunit activation.

Methods. Data presented in the text is restricted to channel-containing patches, defined as those showing Cl^- channel activation by agonists or depolarization. This included 23/46 recordings from normal tracheal cells and 11/22 recordings from CF cells, so that 50% of excised membrane patches contained Cl^- channels capable of activation. This agrees with our previous finding³ that Cl^- channels in 54% of membrane patches were activated by adrenaline or cAMP during cell-attached recording from normal airway cells, and implies that activation in excised membranes assays the same channel population as that evoked by receptor-mediated, cAMP-dependent activation pathways in intact cells. The C-subunit of the cAMP-dependent protein kinase was prepared from bovine heart by the method of Sugden *et al.*¹⁶. Specific activity of the C-subunit preparations ranged from 0.3 to 1.4 units per mg protein, where a unit of kinase catalyses the transfer of $1 \mu\text{mol}$ of ^{32}P from ATP to histone in 1 min^{12} . Kinase preparations from commercial suppliers were not electrophoretically pure and were 500–1,000 times less active than our preparations. A time of 8–10 min was allowed for C-subunit activation of either normal or CF membranes. All normal membranes activated 2–7 min after adding the kinase. Part of this delay may be due to the time required for diffusion of the protein from its site of addition to the membrane, but a similar variation in time for C-subunit-induced channel activation was observed in experiments where patch electrodes were dipped directly into kinase-containing solutions¹⁷.



block in their cAMP-mediated activation lies distal to induction of cAMP-dependent protein kinase activity and focuses our attention on the Cl^- channel and its membrane-associated regulatory proteins as the probable site of the CF defect.

In salt-secreting epithelia, apical-membrane Cl^- permeability is the rate-determining step in secretion and is controlled by a variety of tissue-specific hormones and neurotransmitters whose secretory effects are mediated intracellularly by cAMP⁶. Secretagogues increase Cl^- conductance by stimulating the activity of a 45 pS Cl^- -selective channel that is found in the apical membranes of a variety of secretory epithelial cells, including those of human airways⁷. Although cAMP fails to activate this channel during cell-attached recording from CF airway cells in patch-clamp studies, it becomes active in excised CF membrane patches, where its conduction properties are normal^{3,4}. This activation of Cl^- channels following patch excision was traced to an effect of depolarizing membrane voltages, which were found to activate Cl^- channels in normal and CF membranes in the absence of secretory stimuli. Figure 1 shows this voltage-induced activation. The membrane patch was excised and held at -60 mV for 8 min; during this period, there was no channel activity. Outward Cl^- currents were evoked at 10.5 min, 2.5 min after reversing the holding potential (V_H) to $+60$ mV. In other experiments, channels remained quiet for at least 20 min at negative voltages, but activated 0.25–7 min after clamping excised membranes at depolarizing voltages of $\geq +40$ mV. In contrast to the properties of excitable membrane ion channels⁸, voltage-induced activation of airway cell Cl^- channels was irreversible; channels subsequently remained active over the entire voltage range (Fig. 1). To avoid voltage-induced activation, we maintained excised membranes at physiological voltage (-50 mV)⁹ during exposure to putative regulatory factors in an effort to examine phosphorylation-mediated activation of Cl^- channels in normal and CF membranes¹⁰. Voltage-induced activation, however, provided a convenient assay for the presence of Cl^- channels in patches where regulatory substances were ineffective (see below).

In control experiments, patches excised from normal airway cells were tested for endogenous kinase activity by exposing the cytoplasmic membrane surface to ATP (0.1–1.0 mM, $n = 9$) or ATP plus cAMP (10 μM , $n = 5$). This cAMP concentration

exceeds that necessary for maximal binding to the regulatory subunit of cAMP-dependent protein kinase, as well as that necessary for maximal kinase activity of the holoenzyme *in vitro*¹¹. No Cl^- channel activity was observed during 8–10 min of recording from patches later shown to contain Cl^- channels by either voltage- or C-subunit-induced activation (Fig. 2). Thus, we found no evidence of a non-specific or cAMP-dependent kinase capable of activating Cl^- channels in excised airway-cell membranes.

In the presence of 0.1–1.0 mM ATP, addition of the purified C subunit (0.05–0.5 μM) evoked Cl^- channel activity in 21/23 channel-containing membrane patches from normal cells (Fig. 2). These concentrations of the C subunit agree well with the physiological range of intracellular holoenzyme concentrations (0.2–0.7 μM)¹². The activation evoked by the C subunit appears to be associated with phosphorylation, as membranes exposed to the kinase for 8–10 min in the absence of ATP were quiet, but subsequent addition of ATP activated Cl^- channels within 2–3 min ($n = 6$). Exposure to 0.5 μM bovine serum albumin plus 1 mM ATP for 10 min did not produce activation in membranes subsequently shown to contain Cl^- channels by voltage-induced activation ($n = 3$). In 2/23 experiments on normal cells, the C subunit failed to activate channels that were evoked later by

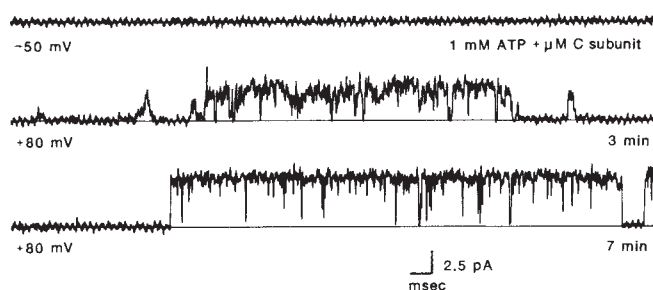


Fig. 3 Failure of C-subunit activation in a CF membrane patch. Top trace: no activity was observed in the presence of ATP plus the C subunit for 10 min; Middle trace: outward Cl^- current was activated 3 min after switching the holding potential to $+80$ mV; Bottom trace: outward Cl^- current 7 min after depolarization.

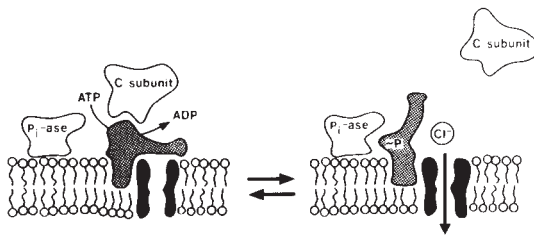


Fig. 4 Schematic representation of the events in Cl^- channel activation by the C subunit. The Cl^- channel is shown as a complex having conduction and regulatory components in the apical-membrane bilayer. P_i -ase, membrane-associated phosphatase. See text for other details.

depolarization. There may be a small population of 'latent' channels in these cells, as membranes excised from normal cells sometimes show channel activity where none could be shown in response to cAMP-dependent agonists in the same patch during on-cell recording^{3,4}. This additional activation is presumably voltage induced.

Membrane patches excised from CF cells uniformly failed to respond to the C subunit in identical protocols. A representative experiment is shown in Fig. 3. In 11 experiments on cells from three CF patients, there was no response to the C subunit, but subsequent depolarization to +80 mV evoked channel activity. As noted previously^{3,4}, the current-voltage relations, kinetics and selectivity of these channels do not differ from normal.

The failure of the C subunit to activate membrane patches derived from CF cells was a uniform finding, but the degree of Cl^- channel activity evoked by the C subunit in normal membranes was variable. The open-channel probability (P_0) of inward Cl^- currents recorded at -50 mV after voltage activation was ~0.2 whereas P_0 for the inward currents stimulated by phosphorylation varied from 0.2 to 0.9. Although the basis of this difference in activity is unknown, some aspect of the kinase- and voltage-induced activations appear to differ. Variations in kinase-induced activity could result from the presence of a membrane-associated phosphatase that competes with the C-subunit effect in excised patches. Consistent with this idea are the results of preliminary studies ($n=3$) showing inactivation of inward Cl^- currents 2-5 min after washing C subunit and ATP from the bath.

Figure 4 shows a simple model that is consistent with our experimental findings and provides a conceptual framework for future studies. It depicts the apical-membrane Cl^- channel as a complex comprised of a conduction pathway and a regulatory component that controls channel activity. In the absence of agonist, the conduction pathway is occluded by the regulatory component, and the channel is inactive. The membrane apparently does not contain a kinase whose activation by cAMP and/or ATP can open the channel. Extrinsic agonists (for example, β -agonists for airway cells) increase intracellular cAMP, freeing the C subunit of the soluble, cAMP-dependent protein kinase. The C subunit phosphorylates the regulatory component of the channel complex, inducing a conformational change that exposes the conduction pathway. A membrane-bound phosphatase could compete with the stimulatory effects of the kinase, and return channels to an inactive state through dephosphorylation when secretory signals dissipate.

According to this model, there are two possibilities that can explain the failure of the C subunit to activate the Cl^- channel complex of CF cells: (1) the regulatory component is not a suitable substrate for phosphorylation, (2) its phosphorylation does not produce the conformational change necessary for conduction. Although the presence of separate components for conduction and regulation are not required to explain the results, other ion-conducting channels appear to be comprised of multiple components¹³⁻¹⁵, and this feature of the model provides a

conceptual basis for the voltage-induced activation. That is, depolarizing voltages might cause the regulatory component to leave the isolated membrane patch, exposing the underlying conduction pathway whose properties are indistinguishable in normal and CF cells.

This work was supported by grants from the NIH and the U.S. Cystic Fibrosis Foundation. We thank J. Ram and J. Brugge for maintaining the cell cultures, E. Walthall for data analysis, D. J. Benos for reading the manuscript, the Tissue Procurement Service at the University of Alabama at Birmingham for normal airway tissues and CF Center staff at the University of Alabama at Birmingham, Baylor College of Medicine and the University of Minnesota for providing the CF airway tissues for these studies.

Received 25 September; accepted 26 November 1987.

1. Quinton, P. M. *Nature* **301**, 421-422 (1983).
2. Knowles, M. R., Gatz, J. T. & Boucher, R. C. *J. clin. Invest.* **71**, 1410-1417 (1983).
3. Frizzell, R. A., Reckemmer, G. & Shoemaker, R. L. *Science* **233**, 558-569 (1986).
4. Welsh, M. J. & Liedtke, C. M. *Nature* **322**, 467-470 (1986).
5. Boucher, R. C., Stutts, M. J., Knowles, M. R., Catley, L. & Gatz, J. T. *J. clin. Invest.* **78**, 1245-1252 (1986).
6. Frizzell, R. A., Field, M. & Schultz, S. G. *Am. J. Physiol.* **236**, F1-F8 (1979).
7. Frizzell, R. A. *Trends Neurosci.* **10**, 190-193 (1987).
8. Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer, Sunderland, Massachusetts, 1984).
9. Welsh, M. J., Smith, P. L. & Frizzell, R. A. *J. Membrane Biol.* **70**, 227-238 (1982).
10. Levitan, I. B. *J. Membrane Biol.* **87**, 177-190 (1985).
11. Schwachheimer, K. & Hofmann, F. *J. biol. Chem.* **252**, 7690-7696 (1977).
12. Flockhart, D. A. & Corbin, J. D. *CRC Rev. Biochem.* **10**, 133-186 (1982).
13. Hartshorne, R. P. & Catterall, W. P. *J. biol. Chem.* **259**, 1667-1675 (1984).
14. Contri-Tronconi, B. M. & Raftery, M. A. *Rev. Biochem.* **51**, 491-530 (1982).
15. Benos, D. J., Saccomani, G. & Sariban-Sohrab, S. *J. biol. Chem.* **262**, 10613-10618 (1987).
16. Sugden, P. H., Holladay, L. A., Reimann, E. M. & Corbin, J. D. *Biochem. J.* **159**, 409-422 (1976).
17. Ewald, D. A., Williams, A. & Levitan, I. B. *Nature* **315**, 503-506 (1985).

Subcellular fractionation of dystrophin to the triads of skeletal muscle

Eric P. Hoffman*, C. Michael Knudson†, Kevin P. Campbell‡ & Louis M. Kunkel*†

* Department of Pediatrics, Harvard Medical School and Division of Genetics, Children's Hospital, and
† Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA

‡ Department of Physiology and Biophysics, University of Iowa, Iowa City, Iowa 52242, USA

Duchenne muscular dystrophy (DMD) is a human X-linked biochemical defect resulting in the progressive wasting of skeletal muscle of affected individuals¹. It is the most common and is considered to be the most devastating of the muscular dystrophies, affecting about 1 in 3,500 live-born males². The gene that, when defective, results in this disorder was recently isolated³⁻⁶. Using the cloned complementary DNA sequences corresponding to the DMD gene, antibodies have been produced that react with a protein species of relative molecular mass (M_r) ~400,000 (400K) which was absent in two DMD-affected individuals and in *mdx* mice⁷. This protein species is called dystrophin because of its identification by molecular-genetic analysis of affected individuals. Here we show that dystrophin is associated with the triadic junctions in skeletal muscle, and is therefore probably involved with Ca^{2+} homeostasis. We also show that the ~450K ryanodine receptor/sarcoplasmic reticulum Ca^{2+} channel⁸, which has the large size and subcellular distribution characteristics of dystrophin, is an immunologically distinct protein species.

Dystrophin messenger RNA^{3,5,6,9} and protein⁷ are present in all types of mature, terminally differentiated muscle tissue.

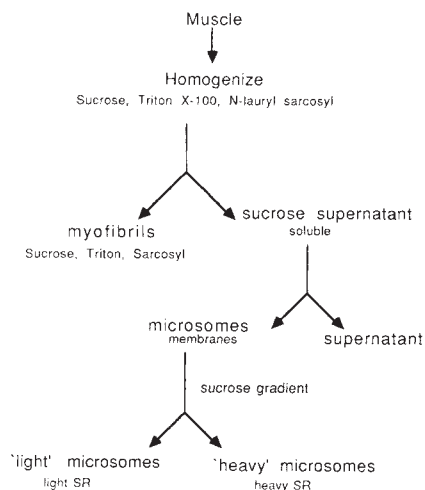


Fig. 1 Schematic diagram of subcellular fractionations, showing the procedure for fractionation of mouse skeletal muscle.

Methods. For detergent-insoluble myofibrils, freshly dissected muscle (2 g) was homogenized in a Waring blender at full speed in 15 ml homogenization buffer (10 mM HEPES pH 7.1, 10 mM EGTA, 1 mM iodoacetamide, 0.8 mM benzamide, $0.5 \mu\text{g ml}^{-1}$ aprotinin, $0.5 \mu\text{g ml}^{-1}$ leupeptin, $0.25 \mu\text{g ml}^{-1}$ pepstatin A, supplemented with solid phenylmethylsulphonyl fluoride) containing either 0.25% Triton X-100 or 0.1% N-lauryl sarcosyl. Insoluble proteins were pelleted at 10,000 r.p.m. in a Beckman JA-13 rotor, pellets resuspended in homogenization buffer, and the protein concentration of the suspension determined using the Bio-Rad protein assay. Samples were diluted with sample buffer and boiled before fractionation (see Fig. 2, lanes 2, 3, 11). Sucrose homogenate fractions (see Fig. 2, lanes 4–9, 12–13) were prepared as described¹¹. Mouse rear-leg skeletal muscle and heart was dissected and immediately homogenized as above, supplemented with 0.25% sucrose. Myofibrillar fractions were isolated by centrifugation as above for 20 min. The skeletal muscle myofibrillar pellet contained two layers (Fig. 2; lanes 4, 5), whereas the heart myofibrillar pellet was homogeneous (Fig. 2; lane 12). The pellets were resuspended in homogenization buffer without sucrose. An aliquot of the low-speed supernatant was reserved (Fig. 2; lane 6) and the remainder was centrifuged at 40,000 r.p.m. in a Beckman SW-40 rotor for 20 min. An aliquot of the pellet was resuspended in homogenization buffer and reserved (microsomes, Fig. 2; lanes 7, 13). The rest of the microsomal pellets were resuspended and fractionated on a discontinuous sucrose gradient for the isolation of light and heavy microsomes (Fig. 2; lanes 8, 9) as described for mouse skeletal muscle¹², except that protease inhibitors were added.

Dystrophin is thought to have a structural role in myofibres because of its high degree of conservation between mice and humans, the 'structural' characteristics of the amino-terminal one-third of the primary amino-acid sequence^{4,5}, and the significant similarity of the amino terminus to the actin-binding site of the cytoskeletal protein α -actinin¹⁰. Consistent with this hypothesis, the protein is generally insoluble in the detergent Triton, but the very low abundance of dystrophin in myofibres ($\sim 0.002\%$ of total muscle protein⁷) indicates that it is probably not a prominent component of the myofibrillar matrix. We sought to determine the subcellular compartment in which dystrophin is localized. We decided to perform fractionation of subcellular components followed by Western analysis rather than immunofluorescent localization because of the artefacts inherent to immunocytochemical techniques when detecting very low abundance proteins with polyclonal antisera.

We dissected skeletal muscle from normal, adult mice and fractionated it into detergent-insoluble (Triton X-100, N-lauryl sarcosyl) and sucrose-homogenized myofibrillar, soluble and microsomal fractions¹¹. We further separated the microsomal fraction into 'light' and 'heavy' microsomes using a sucrose-step

gradient¹² (see Fig. 1). We also performed similar fractionations for mouse heart. We incubated identical Western blots with antibodies directed against dystrophin⁷ (Fig. 2a); the $(\text{Ca}^{2+} + \text{Mg}^{2+})\text{ATPase}$ ¹³ (Fig. 2b); the ryanodine receptor^{8,14} (Fig. 2c); or tropomyosin (Fig. 2d). As expected, tropomyosin and cross-reacting myosin are enriched in the myofibrillar fractions¹⁵ (Fig. 2d; lanes 2–5, 11–12), whereas the $(\text{Ca}^{2+} + \text{Mg}^{2+})\text{ATPase}$ and ryanodine receptor are both enriched in the microsomal fractions (Fig. 2b,c; lanes 7–9, 13)^{12,14,16}. The ATPase monoclonal antibodies used do not cross-react with the cardiac form of this protein (Fig. 2b; lanes 10–13)¹³.

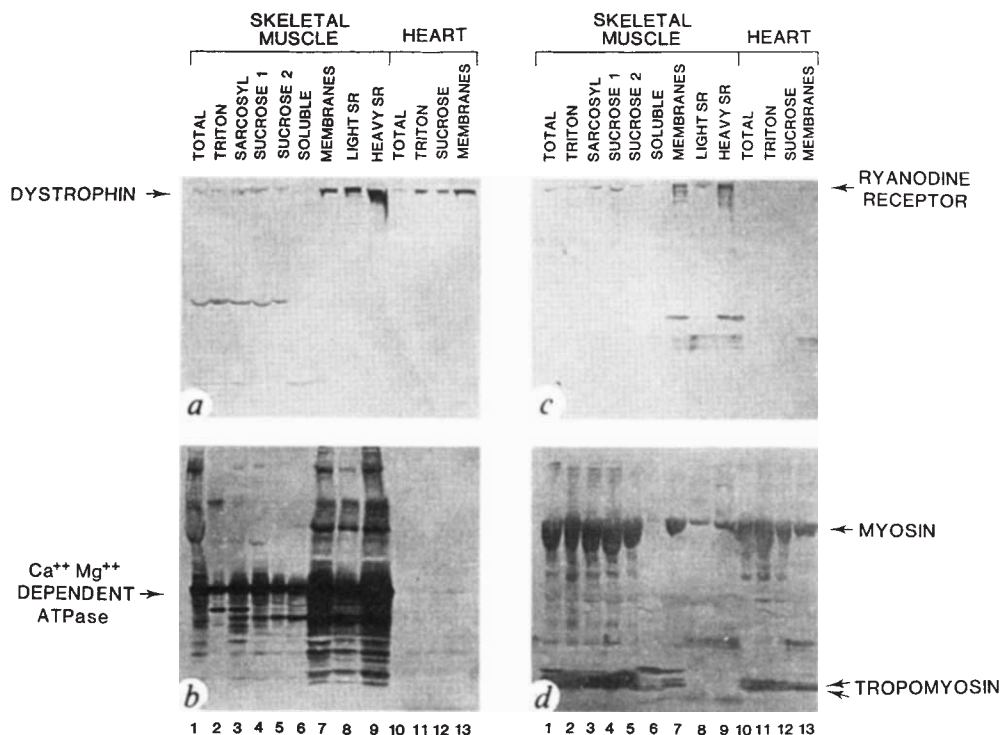
Dystrophin parallels the localization of the ryanodine receptor and $(\text{Ca}^{2+} + \text{Mg}^{2+})\text{ATPase}$, being enriched in the microsomal fractions (Fig. 2a; lanes 7–9, 13), with a particularly strong enrichment in the heavy fraction (Fig. 2a; lane 9). The heavy microsomal fraction was originally developed to enrich sarcoplasmic reticulum vesicles derived from the terminal cisternae of the triad junction. Proteins known to be contained in this fraction are the $(\text{Ca}^{2+} + \text{Mg}^{2+})\text{ATPase}$ ¹⁶, calsequestrin¹⁶ and the ryanodine receptor¹⁴. The enrichment of dystrophin in the same fraction, and the lack of enrichment in other fractions, suggests a similar association with the triad region.

To define further the subcellular compartmentalization of dystrophin within the triad junctional region of skeletal muscle, and to determine whether dystrophin and the ryanodine receptor are the same protein, we performed a second subcellular fractionation series using rabbit skeletal muscle. Triadic fractions can be efficiently isolated from rabbit skeletal muscle using the pyrophosphate variant protocol¹¹. Such fractions have been found to be enriched in high-affinity [³H]PN200-110 (radiolabelled dihydropyridine) binding^{17,18} (a junctional T-system marker) and high-affinity [³H]ryanodine binding¹⁴ (a junctional sarcoplasmic reticulum marker). We collected aliquots from all steps of the purification to demonstrate the co-purification of ryanodine receptor, the dihydropyridine receptor and dystrophin in all fractions. Figure 3a shows the Coomassie blue staining of the fractionated muscle, with the myofibrillar proteins myosin, nebulin and titin indicated, as well as the sarcoplasmic reticulum ryanodine receptor, $(\text{Ca}^{2+} + \text{Mg}^{2+})\text{ATPase}$ and calsequestrin. Western blot analysis of gels identical to Fig. 3a using affinity-purified anti-dystrophin antibodies, monoclonal antibody against the 170K subunit of the dihydropyridine receptor or polyclonal antibodies against the ryanodine receptor indicate that all three detected proteins have a very similar subcellular fractionation pattern (Fig. 3b–d). All three of these proteins were highly concentrated within fractions enriched for triadic structures, [³H]ryanodine binding and [³H]PN200-110 binding (Fig. 3b–d; lanes 4–6). Our additional experiments (not shown) show that extraction of triads with 0.6 M KCl does not remove dystrophin; and immunoblot analysis of heavy sarcoplasmic reticulum vesicles¹⁶ compared with isolated triads¹¹ shows the enrichment of dystrophin and the ryanodine receptor in isolated triads. The co-purification of dystrophin with other, well-characterized triad-associated proteins indicates that dystrophin is tightly associated with the triad structures of skeletal muscle.

To test the relationship between the ryanodine receptor and dystrophin directly, we performed further immunoblot experiments using an affinity-purified ryanodine receptor preparation containing Ca^{2+} -channel activity⁸. We stained isolated triads and the purified ryanodine receptor with polyclonal anti-ryanodine receptor and anti-dystrophin antibodies (Fig. 4b,c). We found that anti-dystrophin antibodies do not react with the affinity-purified 450K ryanodine receptor and we did not detect dystrophin in the purified ryanodine receptor preparation. Also, the polyclonal antibodies to the ryanodine receptor do not react with dystrophin (Fig. 4b; lane 1).

The comparison of the immunostaining (Fig. 4b,c) and amido black staining (Fig. 4a) of the triadic fractions also reveals that the ryanodine receptor is larger and much more abundant than

Fig. 2 Localization of dystrophin to the microsomal fractions of mouse skeletal and cardiac muscle by Western blot analysis of the sub-cellular protein fractions described in Fig. 1. Mouse skeletal muscle: lane 1, total SDS-solubilized (50 μ g) (total); lane 2, Triton X-100-insoluble myofibrils (50 μ g) (triton); lane 3, *N*-lauryl sarcosyl-insoluble myofibrils (50 μ g) (sarcosyl); lane 4, sucrose-insoluble myofibrils (dense fraction) (50 μ g) (sucrose 1); lane 5, sucrose-insoluble myofibrils (50 μ g) (sucrose 2); lane 6, soluble/microsomes (50 μ g) (soluble); lane 7, microsomes (50 μ g) (membranes); lane 8, light microsomes (25 μ g) (light SR); lane 9, heavy microsomes (50 μ g) (heavy SR). Mouse cardiac muscle: lane 10, total SDS-solubilized (50 μ g) (total); lane 11, Triton X-100-insoluble myofibrils (50 μ g) (triton); lane 12, sucrose-insoluble myofibrils (50 μ g) (sucrose); lane 13, microsomes (25 μ g) (membranes). Antibodies used: *a*, anti-mouse dystrophin (affinity-purified sheep anti-60K and rabbit anti-30K polyclonals⁷); *b*, mouse anti-chicken fast ($\text{Ca}^{2+} + \text{Mg}^{2+}$) ATPase monoclonal (sarcoplasmic reticulum Ca^{2+} pump)¹³; *c*, guinea pig anti-rabbit ryanodine receptor polyclonal (sarcoplasmic reticulum Ca^{2+} channel); *d*, rabbit anti-chicken smooth-muscle tropomyosin antisera (myofibrillar marker with cross-reacting myosin shown). Dystrophin

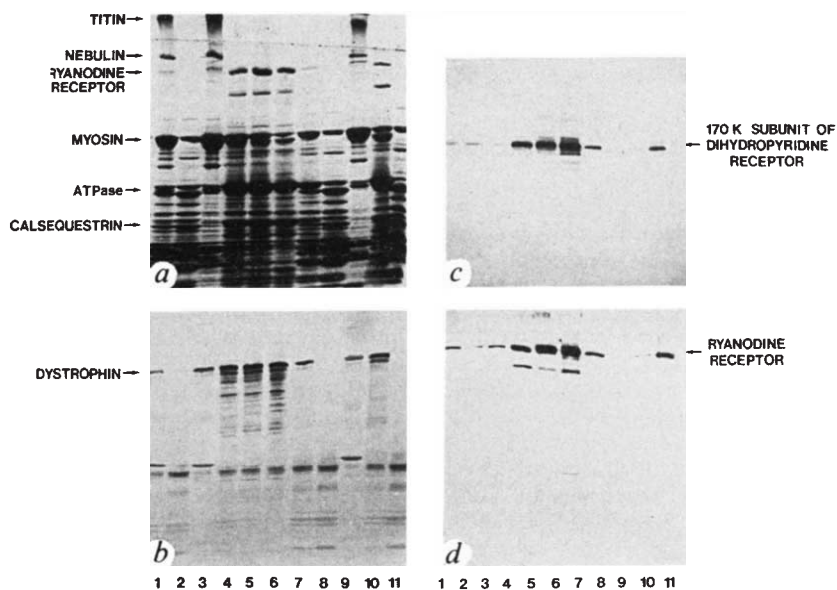


(400K) (*a*) parallels the localization of the sarcoplasmic reticulum markers, the 105K ($\text{Ca}^{2+} + \text{Mg}^{2+}$)ATPase¹³ (*b*) and the 450K ryanodine receptor^{8,14} (*c*). Disassociation of the triads from myofibrils is very inefficient²⁷, explaining the presence of the sarcoplasmic reticulum markers in myofibrillar preparations. Also in *a* is the 100K, skeletal muscle-specific, myofibrillar protein species that cross-reacts with the rabbit anti-30K dystrophin antibodies⁷. The ATPase monoclonal is specific for the fast muscle form of this protein and does not react with the cardiac muscle form (panel *b*, lanes 10–13).

Methods. Total SDS-solubilized muscle (lanes 1, 10). Freshly dissected muscle was minced, boiled in 10 vol sample-loading buffer (10% SDS, 50 mM dithiothreitol, 0.1 M Tris pH 8.0, 10 mM EDTA). SDS-insoluble proteins were precipitated and discarded. All other samples were prepared as described in Fig. 1. Samples were boiled and SDS-insoluble proteins precipitated before gel loading. Protein samples were fractionated on 3.5–12.5% gradient SDS-polyacrylamide gels³⁰, using a 3% stacking gel. Western blotting was as described⁷. Polyclonal antibodies against the purified ryanodine receptor were prepared according to Tung²⁹ by immunization of guinea pigs with gel slices of the purified 450K ryanodine receptor protein.

Fig. 3 Association of dystrophin with triadic junctions of rabbit skeletal muscle. Shown are the following sub-cellular protein fractions from rabbit skeletal muscle. Numbers in parentheses represent [³H]PN200-110 and [³H]ryanodine binding, respectively. Binding (in pmol mg⁻¹) was performed as described^{12,15}. Lane 1, initial muscle homogenate (0.39, 0.09); lane 2, soluble/microsomal fraction of initial homogenate (0.32, 0.12); lane 3, rehomogenized myofibrillar fraction (second homogenate) (0.51, 0.15); lane 4, heavy microsomes from initial homogenate (7.61, 2.53); lane 5, heavy microsomes from second homogenate (10.60, 3.27); lane 6, isolated triads (12.03, 4.25); lane 7, soluble/microsomal fraction from second homogenate (0.46, 1.62); lane 8, soluble/light microsomal fraction from lane 7 (0, 0.05); lane 9, myofibrillar fraction of second homogenate (0.41, 0.11); lane 10, light microsomes from lane 8 (2.50, 0.78); lane 11, soluble fraction (cytosol) (0, 0). *a*, Coomassie blue-stained gel (100 μ g protein per lane); *b*, Western blot using affinity-purified sheep anti-mouse dystrophin antibodies with alkaline phosphatase conjugated donkey anti-sheep IgG secondary. (50 μ g per lane); *c*, Western blot using monoclonal antibody HIF7 directed against the 170K subunit of the dihydropyridine receptor with peroxidase-conjugated goat anti-mouse IgG secondary (50 μ g per lane). *d*, Western blot using polyclonal guinea pig anti-rabbit ryanodine receptor antibodies with peroxidase-conjugated rabbit anti-guinea pig IgG secondary. (50 μ g per lane). In *a*, myofibrillar markers are titin (apparent M_r ~1,500K³¹), nebulin (apparent M_r ~700K³²), myosin (205K) and the sarcoplasmic reticulum marker ($\text{Ca}^{2+} + \text{Mg}^{2+}$)ATPase (105K). Dystrophin (*b*) parallels the localization of the ryanodine receptor (sarcoplasmic reticulum Ca^{2+} channel, *c*), being most highly concentrated in the purified triad fraction (lane 6). Because of differences in the second antibodies used and differences in the sensitivities of the various antisera, the intensity of the bands does not reflect the relative quantity of dystrophin present.

Methods. Total rabbit skeletal muscle was dissected and fractionated into myofibrils, heavy microsomes and triads using a modification¹⁸ of the pyrophosphate-variant procedure¹¹. Light microsomes were isolated as described¹⁶. Protein content of samples was determined³³, and aliquots were solubilized in SDS and fractionated using 3–12% gradient SDS-polyacrylamide gels³⁰ using a 3% stacking gel. Western blot analysis was performed as described¹⁴.



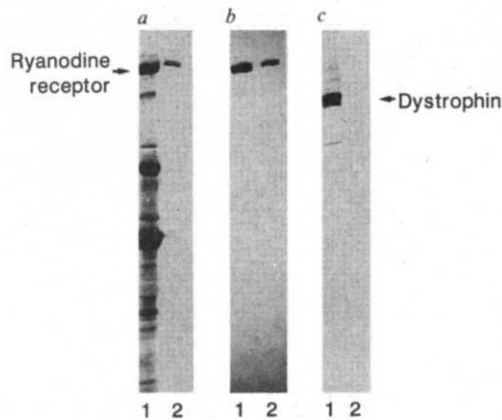


Fig. 4 Dystrophin is unrelated to the ryanodine receptor. Shown are three similar Western blots containing isolated triad structures and purified ryanodine receptor⁸. *a*, Amido black staining. Lane 1, triads (50 µg); lane 2, purified ryanodine receptor (1.5 µg). *b*, Ryanodine receptor seen using guinea pig anti-rabbit ryanodine receptor antibodies with peroxidase-conjugated rabbit anti-guinea pig IgG secondary antibodies. Lane 1, triads (10 µg); lane 2, purified ryanodine receptor (0.7 µg). *c*, Dystrophin seen using sheep anti-mouse dystrophin antibodies with alkaline phosphatase-conjugated donkey anti-sheep IgG. Lane 1, triads (50 µg); lane 2, purified ryanodine receptor (1.5 µg).

dystrophin. The amido black-stained protein species migrating immediately below the abundant ryanodine receptor (Fig. 4*a*; see also Fig. 3*a*) could potentially represent dystrophin. But there are differences between the fractionation of this amido black-stained protein species and the fractionation of dystrophin; the stained protein is not evident in lane 9 of Fig. 3*a*, whereas dystrophin is present in this lane (Fig. 3*b*). Furthermore, in contrast to dystrophin (Fig. 3*b*), the stained protein shows a substantial enrichment in the light microsomal fraction (Fig. 3*a*; lane 10) relative to the heavy microsomal fractions (Fig. 3*a*; lanes 4–6). Dystrophin either co-migrates with this amido black-stained protein species, or is not abundant enough to be stained at all. In either case, dystrophin presumably is a very minor component of the triads. This observation is consistent with previous estimates of the very low relative abundance of dystrophin in striated muscle^{5,7}. The differences in size and abundance between the ryanodine receptor and dystrophin indicate that they are indeed distinct, immunologically unrelated proteins.

The triadic structures of muscle consist of transverse sarcolemmal invaginations (t-tubules) interdigitated between two terminal cisternae of the longitudinally oriented sarcoplasmic reticulum (junctional sarcoplasmic reticulum). These structures occur at the A-I junction of every half sarcomere in higher vertebrates and are responsible for the signal transduction of membrane depolarization into sarcoplasmic reticulum Ca^{2+} release, resulting in myofibrillar contraction (excitation-contraction coupling). Membrane depolarizations triggered by motorneurons are transmitted to the internal portions of myofibres via the t-tubule system, where voltage sensors in the triadic structures (possibly the dihydropyridine receptor of the t-tubules^{19–21}) are thought to initiate Ca^{2+} release from the sarcoplasmic reticulum through the sarcoplasmic reticulum Ca^{2+} channel (the ryanodine receptor)^{8,22}, resulting in myofibril contraction. The junctional sarcoplasmic reticulum is also responsible, together with the non-junctional sarcoplasmic reticulum, for the active uptake and storage of intracellular Ca^{2+} , which is accomplished via the very abundant intrinsic (Ca^{2+} + Mg^{2+})ATPase and calsequestrin, the extrinsic Ca^{2+} -binding protein.

Here we have shown that the protein product of the DMD locus, dystrophin, is a protein associated with triadic structures

in myofibres. Because these structures are essential for the maintenance of calcium homeostasis within the muscle fibre, one can speculate that the absence of one of its potential components, dystrophin, would disrupt this process. It has been reported²³ that increased intracellular Ca^{2+} levels could initiate myofibre death by the activation of phospholipase A, resulting in sarcolemmal dissolution. It is possible that altered Ca^{2+} homeostasis resulting from dystrophin deficiency leads to such a Ca^{2+} -mediated phospholipase activation. This process could explain the very high levels of soluble sarcoplasmic enzymes in the serum of the DMD-affected individual²³.

The amino-terminal 200 amino acids of dystrophin exhibit striking similarity to the actin filament-binding domain of chicken cytoskeletal α -actinin¹⁰. This actin filament-binding domain is highly conserved between the widely divergent *Dictyostelium* and chicken α -actinin, presumably due to the highly conserved nature of the actin filaments to which both proteins bind²⁴. The similarity of the amino terminus of dystrophin to this same conserved region suggests that this similarity represents a functional homology. It is therefore probable that dystrophin interacts with actin filaments in myofibres. We have shown that dystrophin is associated with the triadic structures of muscle. It is tempting to suggest that dystrophin might serve as an anchor for the triads to the myofibrillar cytoskeleton through the binding of actin filaments at its amino terminus. Supporting this hypothesis, DMD-affected muscle (which lacks dystrophin) has been seen to contain poorly aligned triads²⁵ and disorganized t-tubule system networks²⁶.

Clearly there are many other hypothetical roles for dystrophin in the triads. The localization of dystrophin to the triadic junctions, however, implies that Ca^{2+} homeostasis is the most likely process disrupted in DMD. The precise function of dystrophin should soon be elucidated, as its subcellular localization is refined by immunoelectron microscopy and production of monoclonal antibodies, and as its biochemistry is defined by more extensive purifications and binding studies.

We thank Zaven Kaprielian and Douglas Fambrough for (Ca^{2+} + Mg^{2+})ATPase antibodies; Toshiaki Imagawa and Albert Leung for the anti-ryanodine receptor antibody and anti-dihydropyridine receptor antibody, respectively; Michel Koenig for discussions; Angelika Noegel and Walter Witke for communication of *Dictyostelium* α -actinin data; and Steven Kahl for technical assistance. Supported by grants from NIH (L.M.K. and K.P.C.) and MDA (L.M.K. and K.P.C.). E.P.H. is the Harry Zimmerman postdoctoral Fellow of the MDA; K.P.C. is an established investigator of the AHA; L.M.K. is an associate investigator of the Howard Hughes Medical Institute.

Received 6 November; accepted 17 November 1987.

- Walton, J. N. & Gardner-Medwin, D. in *Disorders of Voluntary Muscle* (ed. Walton, S. J.) 481–524 (Churchill Livingstone, Edinburgh, 1981).
- Moser, H. *Hum. Genet.* **66**, 17–40 (1984).
- Monaco, A. P. *et al. Nature* **323**, 646–650 (1986).
- Koenig, M. *et al. Cell* **50**, 509–517 (1987).
- Hoffman, E. P., Monaco, A. P., Feener, C. C. & Kunkel, L. M. *Science* **238**, 347–350 (1987).
- Burghes, A. H. *et al. Nature* **328**, 434–437 (1987).
- Hoffman, E. P., Brown, R. H. & Kunkel, L. M. *Cell* (in the press).
- Imagawa, T., Smith, J. S., Coronado, R. & Campbell, K. P. *J. biol. Chem.* **262**, 16636–16643 (1987).
- Lev, A., Feener, C., Kunkel, L. M. & Brown, R. H. *J. biol. Chem.* (in the press).
- Hammonds, R. G. *Cell* **51**, 1 (1987).
- Mitchell, R. D., Palade, P. & Fleischer, S. *J. Cell Biol.* **96**, 1008–1016 (1983).
- Mrak, R. E. & Fleischer, S. *Muscle Nerve* **5**, 143–151 (1982).
- Kaprielian, Z. & Fambrough, D. M. *Dev. Biol.* (in the press).
- Campbell, K. P. *et al. J. biol. Chem.* **262**, 6460–6463 (1987).
- Stone, D. & Smillie, L. B. *J. biol. Chem.* **253**, 1137–1148 (1978).
- Campbell, K. P., Franzini-Armstrong, C. & Shamo, A. E. *Biochim. biophys. Acta* **602**, 97–116 (1980).
- Leung, A. T., Imagawa, T. & Campbell, K. P. *J. biol. Chem.* **253**, 1137–1148 (1978).
- Sharp, A. H., Imagawa, T., Leung, A. T. & Campbell, K. P. *J. biol. Chem.* **262**, 12309–12315 (1987).
- Bios, E. & Brum, G. *Nature* **325**, 717–720 (1987).
- Tanabe, T. *et al. Nature* **328**, 313–318 (1987).
- Leung, A. T., Imagawa, T., Block, B., Franzini-Armstrong, C. & Campbell, K. P. *J. biol. Chem.* (in the press).
- Smith, J. S., Coronado, R. & Meissner, G. *Nature* **316**, 446–449 (1985).

23. Jackson, M. J., Jones, D. A. & Edwards, R. H. *Eur. J. clin. Invest.* **13**, 369-374 (1984).
24. Noegel, A., Witke, W. & Schleicher, M. *Fedn Eur. Biochem. Sci.* **221**, 391-396 (1987).
25. Watkins, S. C. & Cullen, M. J. *J. neurol. Sci.* (in the press).
26. Oguchi, K., Yanagisawa, N. & Tsukagoshi, H. *J. neurol. Sci.* **57**, 333-341 (1982).
27. Rodemann, H. P., Waxman, L. & Goldberg, A. L. *J. biol. Chem.* **257**, 8716-8723 (1982).
28. Martonosi, A. N. in *Myology* (eds Engel, A. G. & Banker, B. Q.) 521-562 (McGraw Hill, New York, 1986).
29. Tung, A. S. *Meth. Enzym.* **93**, 12-23 (1983).
30. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
31. Wang, K. *Cell Muscle Motil.* **6**, 314-369 (1985).
32. Sugita, H. *et al. Proc. Jap. Acad.* **63**, 107-110 (1987).
33. Peterson, G. L. *Analyt. Biochem.* **83**, 346-356 (1977).

Identification of receptor contact site involved in receptor-G protein coupling

Kathleen A. Sullivan, R. Tyler Miller,
Susan B. Masters, Barry Beiderman, Warren Heideman
& Henry R. Bourne*

Departments of Pharmacology and Medicine and the Cardiovascular Research Institute, University of California, San Francisco, California 94143-0450, USA

The mammalian G proteins transduce information from extracellular signals, including neurotransmitters, hormones and sensory stimuli, into regulation of effector enzymes or ion channels within cells. Triggered by appropriate extracellular signals, receptor proteins specifically activate members of the G protein family by catalysing replacement of GDP by GTP at the guanine nucleotide binding site. Like the receptor proteins¹, the heterotrimeric G proteins exhibit impressive structural similarities^{2,3}, suggesting that all receptor-G protein interactions use homologous structural elements and a single molecular mechanism. Topologically equivalent portions of each G protein may therefore interact with the appropriate receptor. We recently predicted the secondary structure of a composite G protein α -chain and proposed that a predicted amphipathic α -helix at the extreme carboxy-terminus of the polypeptide directly contacts receptors⁴. This proposal has now been confirmed by sequencing complementary DNAs of the gene that encodes the α -chain (α_s) of the stimulatory regulator (G_s) of adenylyl cyclase in wild-type cells and in a mutant mouse S49 lymphoma cell line, *unc*, in which G_s cannot be activated by hormone receptors⁵. The sequences reveal a point mutation in the *unc* gene that substitutes a proline residue for an arginine near the carboxy-terminus of the α_s -polypeptide. Expression of recombinant α_{s-unc} in genetically α_s -deficient S49 cells reproduces the *unc* phenotype.

Adenylyl cyclase in S49 *unc* mutant cells^{5,6} responds normally to agents that act directly on G_s , such as cholera toxin, A1F₄ ion, and hydrolysis-resistant guanine nucleotides. But adenylyl cyclase in *unc* cells cannot be stimulated by hormonal agonists that act on β -adrenoceptors or prostaglandin E₁ receptors. The inability of guanine nucleotides to regulate affinity of agonist binding by β -adrenoceptors in *unc* mutant membranes also indicates that the mutation may impair communication between G_s and receptors. Three lines of evidence indicate that the *unc* mutation is in the gene encoding α_s : (1) Addition of pure G_s to *unc* mutant membranes restores both responsiveness of adenylyl cyclase to hormones and sensitivity of agonist binding by β -adrenoceptors to regulation by guanine nucleotides⁷. (2) In keeping with the idea that the *unc* mutation results in substitution of an amino acid, the isoelectric point of α_s in mutant *unc* cells is more acidic than that of the wild type protein, by approximately one electrical charge⁸. (3) S49 cells bearing the *cyc*⁻ mutation, which abolishes expression of α_s ⁹, fail to complement the defect of *unc* mutant cells in somatic hybrids⁶, suggesting

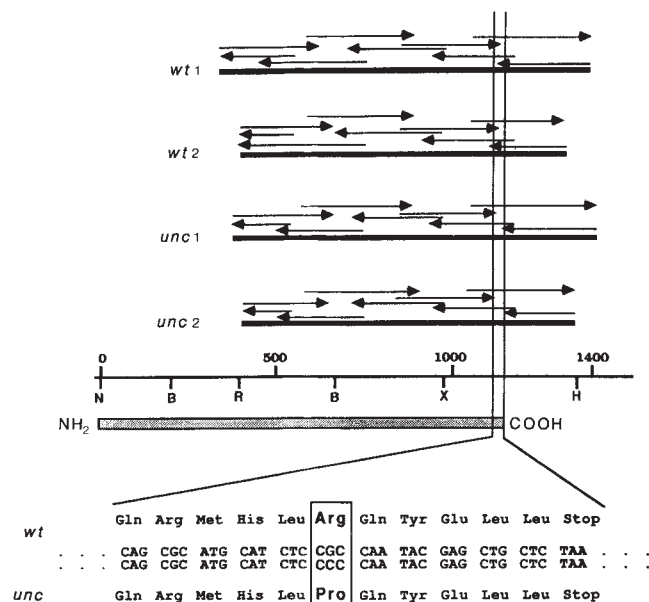


Fig. 1 The *unc* mutation predicts replacement of an arginine by a proline residue at position 372 of the α_s -polypeptide. The shaded bar represents the 1,131-base coding region of a previously reported¹⁰ mouse α_s cDNA. Solid bars (top) represent the lengths of two wild type (wt1, wt2) and two *unc* (unc1, unc2) cDNAs; arrows show lengths, positions, and directions of the sequencing reactions. Numbers on the horizontal line indicate bases in the mouse α_s cDNA¹⁰; letters indicate restriction-endonuclease sites (N, *Nco*I; B, *Bam*HI; X, *Xma*I; H, *Hind*III). DNA and predicted amino-acid sequences of wild-type and *unc* carboxy-termini (bottom) reveal substitution of guanosine by cytosine at base 1,115, converting an arginine to a proline codon at amino-acid residue 372. All four cDNAs showed two differences from the previously reported¹⁰ α_s base sequence, at base positions 639 (adenosine to guanosine) and 933 (cytosine to guanosine); each of these is a conservative substitution in the third base of a redundant codon, producing no change in amino-acid sequence.

Methods. *Unc* mutant cells were propagated, as previously described⁶, in roller bottles containing Dulbecco's modified Eagle medium (DMEM) with 4.5 g per l glucose and 10% heat-inactivated horse serum. Before preparation of cDNA libraries, adenylyl cyclase phenotypes of the cells were verified by enzyme assays of membrane extracts. Poly(A)⁺ RNA (3 μ g) from each cell line was used to prepare cDNA libraries, as previously described¹¹, using a cDNA synthesis kit (Amersham). The cDNA was treated with *Eco*RI methylase and ligated into λ gt10 phage at an *Eco*RI site. Libraries were screened with a 400 base-pair (bp) (nucleotides 383-782) DNA probe derived from a mouse α_s cDNA¹⁰. cDNA inserts selected in this manner were subcloned into M13 vectors and sequenced using the dideoxy technique^{12,13}.

that the *unc* and *cyc*⁻ mutations are located in the same gene.

We prepared cDNA libraries from S49 wild type and *unc* mutant cells and isolated multiple partial cDNAs encoding α_s from each (see Fig. 1 legend). Two independent wild type cDNA inserts, encoding ~70% of the α_s open reading frame, had nucleotide sequences identical to that of a reported¹⁰ mouse α_s cDNA, except for two conservative base substitutions that did not alter the predicted amino acid sequence. Two α_s cDNAs from the *unc* mutant gene showed a single additional mutation, a replacement of guanosine by cytosine at base 1,115 (Fig. 1). This would replace arginine by proline at amino acid position 372, six residues from the polypeptide's carboxy-terminus; the predicted loss of a positive charge accounts for the altered isoelectric point of *unc* α_s (ref. 8).

Because the *unc* cDNA sequence was incomplete (Fig. 1), it was necessary to prove that the mutation at base 1,115 causes the *unc* phenotype. We constructed a chimaeric α_s DNA

* To whom correspondence should be addressed at the Department of Pharmacology.

Fig. 2 Phenotypes of *wild type* (*wt*), *unc*, and *cyc⁻kin⁻* S49 cells compared with those of *cyc⁻* cells expressing recombinant rat α_s (*csw.3*) or α_{s-unc} (*ksu.1*). **a**, Cholera toxin substrates: particulate extracts of sonicated cells¹⁵ were radiolabelled by cholera toxin-catalysed ADP-ribosylation (20 μ g ml⁻¹ activated toxin, 50 μ M [³²P]NAD⁺), as described previously¹⁵. Extracts (representing 1.25×10^7 cells per lane) were then subjected to SDS-polyacrylamide gel electrophoresis as described¹⁵, and subjected to autoradiography. Cell types include: wild type (lane 1); *csw.3* (*cyc⁻* expressing normal rat α_s , lane 2); *unc* (lane 3); *ksu.1* (*cyc⁻kin⁻* expressing α_{s-unc} , lane 4); *cyc⁻kin⁻* (lane 5). Numbers (left) indicate size (K) of standard proteins, and arrows (right) indicate positions of the 52K and 45K cholera toxin substrates. **b**, Adenylyl cyclase activities. Particulate extracts (1×10^7 cells per assay) of sonicated cells were assayed for adenylyl cyclase activity, as described previously¹⁵, in 100 μ l of 50 mM Tris buffer, pH 8, 2.5 mM MgCl₂, 1 mM EDTA, 2 mM β -mercaptoethanol, 0.1 mg ml⁻¹ bovine serum albumin, 10 mM creatine phosphate, 10 U ml⁻¹ creatine phosphokinase, 1 mM cyclic AMP, 0.4 mM [³²P]ATP, and other reagents as indicated (GTP, 50 μ M; isoproterenol [Iso], 100 μ M; guanylyl-5'-imidodiphosphate [Gpp(NH)p], 100 μ M; NaF, 10 mM).

Methods. Wild type, *cyc⁻* and *unc* cells have been described previously^{5,6}. The *cyc⁻kin⁻* double mutant was isolated by subjecting a population of *cyc⁻* cells to selection in soft agar containing 3 mM N⁶-2'-O-dibutyryl cyclic AMP, as described⁶; the adenylyl cyclase phenotype of this double mutant was identical to that of *cyc⁻*. A full length rat cDNA encoding the 52K α_s -polypeptide¹⁴ was introduced into the retroviral vector, pMV7 (ref. 16), using the unique *Hind*III restriction endonuclease site. α_{s-unc} DNA was prepared by digesting both the rat α_s cDNA and the mouse *unc*1 cDNA with *Xma*I and *Hind*III; the 4,332-bp rat α_s DNA fragment and a 363-bp fragment from *unc*1 were recovered from a low melt agarose gel, ligated, and introduced into the *Hind*III site of the pMV7 vector. Transfection: cells (5×10^5 per 100 mm plate) of the PA12 packaging cell line¹⁷ were transfected by calcium phosphate co-precipitation with 15 μ g high molecular weight carrier DNA and 15 μ g pMV7 DNA with the appropriate α_s insert, as described¹⁸. Medium (DMEM with 3 g l⁻¹ glucose and 10% fetal bovine serum) was replaced 16 h later; at 36 h the medium was removed and filtered through a 0.2 μ filter (Nalgene). ψ 2 cells¹⁹ (5×10^5) were infected by 16 h incubation with 5 ml of the filtrate plus 5 ml medium and 8 μ g polybrene ml⁻¹. Infected ψ 2 cells were selected by growth in the presence of G418 (400 μ g per ml) for 10–14 days. Before infection of S49 cells, the ψ 2 medium was replaced for 24 h with fresh medium lacking G418. S49 *cyc⁻* cells (or *cyc⁻kin⁻* cells) were infected by co-culturing overnight with a layer of G418-resistant ψ 2 cells. S49 cells were recovered by aspiration, washed once with fresh medium (DMEM with 4.5 g l⁻¹ glucose and 10% heat-inactivated horse serum), and cultured for 24 h before cloning in soft agar, as described⁶, in the presence of 400 μ g ml⁻¹ G418.

sequence (α_{s-unc}) in which the *Xma*I–*Hind*III fragment of the *unc* cDNA, containing the mutation of interest (Fig. 1), was substituted for the corresponding fragment in a rat cDNA¹⁴ encoding the complete α_s polypeptide (relative molecular mass, M_r , 52 K); the chimaeric sequence encodes a polypeptide that differs from the normal rat α_s only in the replacement of arginine by proline at a position six residues from the carboxy-terminus. Using a retroviral vector, normal rat α_s and α_{s-unc} were expressed in S49 *cyc⁻* cells (see Fig. 2 legend). As expected, expression of both DNAs produced a 52K α_s polypeptide, detected by cholera toxin-catalysed ADP-ribosylation (Fig. 2a). Adenylyl cyclase activity in extracts of cells expressing normal rat α_s (*csw.3* cells) resembled that of wild-type S49 cells, whereas cells expressing α_{s-unc} (*ksu.1* cells) showed the *unc* phenotype (Fig. 2b), in which a GTP analogue and AlF₄⁻ ion stimulate cyclic AMP synthesis, but isoproterenol, a β -adrenoceptor agonist, does not. This result conclusively proves that the arginine to proline mutation causes the *unc* phenotype.

The covalent modification catalysed by pertussis toxin (PT)

produces a functional lesion in other G proteins that closely resembles the defect of G_s in *unc* mutants. These other G proteins^{2,3} include G_i, the inhibitory regulator of adenylyl cyclase, G_o, an abundant brain G protein of unknown function and transducin, the GTP-binding protein that mediates photo-transduction in retinal rod cells. PT catalyses ADP-ribosylation of a conserved cysteine residue near the carboxy-terminus of the α -chain of these proteins (Fig. 3), in each case producing a G protein whose ability to interact with receptors is specifically impaired^{2,3}. PT-catalysed modification of G_i, for example, attenuates hormonal inhibition of adenylyl cyclase²⁰ and diminishes the ability of G_i to regulate the affinity of receptors for binding agonists²¹. Similarly, PT-modified transducin shows reduced affinity for binding its receptor, rhodopsin, and consequently cannot transmit photon signals from rhodopsin to the effector enzyme, cyclic GMP-phosphodiesterase²².

The *unc* mutation of α_s and the ADP-ribose attached by PT to other α -chains cause equivalent defects in G protein function. This equivalence indicates that the carboxy-termini of all the G

Fig. 3 The carboxy-terminal 20 amino acids of the α -chains of G_s, G_i, G_o and transducin⁴, compared to their respective abilities to couple functionally with rhodopsin (Rho), β -adrenoceptors (β), and α_2 -adrenoceptors (α_2). Amino-acid residues enclosed by heavy lines include the arginine residue of α_s that is mutated in *unc* and the cysteine residue ADP-ribosylated by PT in G_i, G_o, and transducin. The bottom line shows an internal sequence (residues 207–226)¹⁴ of a retinal protein called arrestin, or 48K; amino-acid residues that are identical or conserved with respect to the transducin α chain carboxy-terminus, are designated by rectangles or ovals, respectively.

		Rho	β	α_2
S	Val Phe Asn Asp Cys Arg Asp Ile Ile Gln Arg Met His Leu Arg Gln Tyr Glu Leu Leu	-	+	-
i	Val Phe Asp Ala Val Thr Asp Val Ile Ile Lys Asn Asn Leu Lys Asp Cys Gly Leu Phe	+	-	+
o	Val Phe Asp Ala Val Thr Asp Ile Ile Ile Ala Asn Asn Leu Arg Gly Cys Gly Leu Tyr	+	-	+
t	Val Phe Asp Ala Val Thr Asp Ile Ile Ile Lys Glu Asn Leu Lys Asp Cys Gly Leu Phe	+	-	-
48K	Asp Glu Asn Phe Val Phe Glu Glu Phe Ala Arg Glu Asn Leu Lys Asp Ala Gly Glu Tyr	+	?	?

protein α -chains are critical for binding to receptors, even though the amino acid sequence of α_s appears only distantly related to that of the other α -chains in this region (Fig. 3). This conclusion is strongly reinforced by the recently reported²³ amino-acid sequence of the 48K protein of retinal rod cells. This protein, also called arrestin or S-antigen, competes with transducin for binding to photorhodopsin. An internal sequence of the 48K protein closely resembles that of the carboxy-terminus of α_t ²³; the similarity (Fig. 3) makes it highly probable that direct contacts between this region of α_t and photorhodopsin participate in their binding interaction.

If α -chain carboxyl-termini serve the same function in all G protein α -chains, they should share the same secondary structure and three-dimensional location in the folded G protein. On the basis of their amino acid sequences, we predicted⁴ an amphipathic α -helical structure for these regions of the α chains. It is possible that the arginine 372 to proline substitution in *unc* mutant cells uncouples α_s from receptors by introducing a kink in the presumptive α -helix. Alternatively, the *unc* phenotype could result from the loss of a positively-charged amino acid side chain at this position; all the known α -chains, as well as the retinal 48K protein, have either an arginine or a lysine residue at this position, suggesting that the positive charge may be functionally important.

Variations in the amino acid sequences of G protein regions that bind receptors should allow the different G proteins to discriminate specifically among subsets of receptors. The abilities of rhodopsin and β -adrenoceptors to discriminate among the G proteins²⁴ correlate nicely with the carboxy-terminal sequences of α -chains (Fig. 3). Rhodopsin can stimulate GTP hydrolysis by G_i , G_o and transducin, but not by G_s , in keeping with the nearly identical carboxy-termini of α_i , α_o , and α_t , but not of α_s ; conversely, β -adrenoceptors can stimulate the GTPase activity of G_s much more effectively than that of the other G proteins^{24,25}.

The α -chain carboxy-terminus is not the only region that participates in interactions with receptors, as shown by the observation that the agonist-stimulated α_2 -adrenoceptor can stimulate GTP hydrolysis by G_i and G_o , but not by transducin²⁶. Thus the α_2 receptor can discriminate among G protein α -polypeptides whose carboxy-termini are virtually identical (Fig. 3). This indicates that the specificity of G protein-receptor interactions requires at least one more point of contact between the two signal transducing elements, in addition to the carboxy-terminus of the G protein α -chain.

We thank P. Kirschmeier for providing the pMV7 vector, R. Reed for providing the rat α_s cDNA, Dr A. Gilman for the *unc* cell line and D. Littman, H. Varmus and R. Miesfeld for useful advice. This work was supported by the NIH and the March of Dimes.

Received 5 October; accepted 22 October 1987.

- Dixon, R. A. F. *et al. Nature* **326**, 73-77 (1987).
- Stryer, L. & Bourne, H. R. *A. Rev. Cell Biol.* **2**, 391-419 (1986).
- Gilman, A. G. *A. Rev. Biochem.* **56**, 615-649 (1987).
- Masters, S. B., Stroud, R. M. & Bourne, H. R. *Protein Engng* **1**, 47-54 (1986).
- Haga, T., Ross, E. M., Anderson, H. J. & Gilman, A. G. *Proc. natn. Acad. Sci. U.S.A.* **74**, 2016-2020 (1977).
- Bourne, H. R., Beiderman, B., Steinberg, F. & Brothers, V. M. *Molec. Pharmac.* **22**, 204-210 (1982).
- Sternweis, P. C. & Gilman, A. G. *J. biol. Chem.* **254**, 3333-3340 (1979).
- Schleifer, L. S., Garrison, J. C., Sternweis, P. C., Northup, J. K. & Gilman, A. G. *J. biol. Chem.* **255**, 2641-2644 (1980).
- Harris, B. A., Robishaw, J. D., Mumby, S. M. & Gilman, A. G. *Science* **229**, 1274-1277 (1984).
- Sullivan, K. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 6687-6691 (1986).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (eds) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor, New York, 1982).
- Sanger, F., Nicklen, F. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Messing, J., Crea, R. & Seeburg, P. *Nucleic Acid Res.* **9**, 309-321 (1981).
- Jones, D. T. & Reed, R. R. *J. biol. Chem.*, (in the press).
- Salomon, M. R. & Bourne, H. R. *Molec. Pharmac.* **19**, 109-116 (1981).
- Maddon, P. J. *et al. Cell* **42**, 93-104 (1985).
- Miller, A. D., Law, M. F. & Verma, I. M. *Molec. Cell Biol.* **5**, 431-437 (1985).
- Gorman, C., in *DNA Cloning: Vol. II. A Practical Approach* (ed. Glover, D. M.) 143-190 (IRL, Oxford, 1985).
- Mann, R., Mulligan, R. C. & Baltimore, D. *Cell* **33**, 153-159 (1983).

- Katada, T., Amano, T. & Ui, M. *J. biol. Chem.* **257**, 3739-3746 (1982).
- Kurose, M., Katada, T., Amano, T. & Ui, M. *J. biol. Chem.* **258**, 4870-4875 (1983).
- Van Dop, C. *et al. J. biol. Chem.* **259**, 23-26 (1984).
- Wistow, G. J., Katial, A., Craft, C. & Shinohara, T. *FEBS Lett.* **196**, 23-28 (1986).
- Cerione, R. A. *et al. J. biol. Chem.* **260**, 1493-1500 (1985).
- Asano, T., Katada, T., Gilman, A. G. & Ross, E. M. *J. biol. Chem.* **259**, 9351-9354 (1984).
- Cerione, R. A. *et al. J. biol. Chem.* **261**, 3901-3909 (1986).

Activation of a G protein promotes agonist responses to calcium channel ligands

Roderick H. Scott & Annette C. Dolphin*

Department of Pharmacology, St George's Hospital Medical School, Cranmer Terrace, London SW17 0RE, UK

The activation of a guanine nucleotide binding (G) protein is an essential step in coupling certain receptors to the inhibition of voltage-activated calcium channels¹⁻³. We have previously observed that analogues of GTP potentiate the effect of receptor agonists and inhibit calcium currents in cultured dorsal root ganglion (DRG) neurones^{2,4}. A residual sustained 'L-type' component⁵ of the calcium channel current is resistant to inhibition by internal guanosine 5'-O-3-thiotriphosphate (GTP- γ -S)⁴. Because calcium channel antagonists such as D600, nifedipine and diltiazem^{6,7} inhibit L currents^{5,8,9}, we examined their effect on GTP- γ -S-modified currents. These compounds all produced a rapid and very marked potentiation of calcium channel currents in the presence of internal GTP- γ -S and this effect was prevented by pertussis toxin which ADP-ribosylates the G proteins G_i/G_o (for review see ref. 10). We suggest that this potentiation indicates that activated G protein can interact with the calcium channel, and that this enhances the action of calcium channel ligands at their agonist sites on the channel in its resting state^{11,12}. These results represent the first electrophysiological evidence that guanine nucleotides are able to influence cellular responses to calcium channel ligands.

We have examined the effect on calcium currents, recorded in the presence of GTP- γ -S, of the phenylalkylamine D600, and nifedipine, a non-chiral 1,4-dihydropyridine, the latter class selectively inhibiting L channels^{5,13}. Under control conditions, D600 produced an immediate small potentiation of the calcium channel current in 3/5 cells, followed by a use-dependent inhibition (Fig. 1a). With GTP- γ -S in the patch pipette, D600 induced an immediate 4.3 $\pm$ 1.2-fold (mean $\pm$ s.e.m.; $n = 6$) potentiation of the maximum current (Fig. 1b, c). This gradually diminished with continued application of D600 for up to 10 min, although the current remained larger than that before application of D600 (Fig. 1b, c). Pre-incubation of DRGs with pertussis toxin, which ADP-ribosylates the α subunits of G_i/G_o , abolished the D600-induced potentiation in the presence of GTP- γ -S (Fig. 1c), and also prevented any transient potentiation by D600 in control cells, but did not affect the inhibition ($n = 5$). No change in the voltage dependence of activation of the calcium channel currents due to D600 was observed, either in control cells ($n = 3$, data not shown) or in the presence of internal GTP- γ -S ($n = 4$; Fig. 1d).

Similar results were obtained using either 30 μ M diltiazem or 5 μ M nifedipine although the response to nifedipine was slower. Under control conditions, with a holding potential (V_H) of -80 mV, nifedipine increased the maximum calcium channel current by 28 $\pm$ 15% ($n = 5$) after 2 min. After 5 min, a plateau of 50 $\pm$ 12% ($n = 5$) inhibition was reached. In contrast the maximum increase in current after 2 min using internal GTP- γ -S was 177 $\pm$ 84% ($P < 0.05$ compared to before nifedipine; $n = 8$). The potentiation then declined, but after 5 min the current was

* To whom correspondence should be addressed.

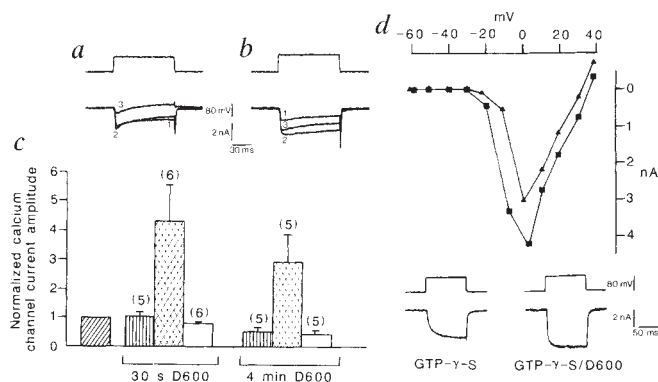


Fig. 1 The effect of internal GTP- γ -S and pertussis toxin on the response to D600. Maximum calcium channel currents were recorded at 0.03 Hz from a holding potential (V_H) of -80 mV to a clamp potential V_C of 0 mV (a) or +10 mV (b) under control conditions (a) and with 500 μ M GTP- γ -S in the patch pipette (b). The current before D600 application is labelled 1. Although it shows a component of inactivation during the 100 ms voltage step in the control (a1) it is non-inactivating in the presence of GTP- γ -S (b1). The first current activated 30 s after the start of application of D600 (10 μ M) is labelled 2. In the control the peak current is slightly potentiated by D600, and there is a greater degree of inactivation (a2). In the presence of internal GTP- γ -S there is a large potentiation of the calcium channel current (b2). The current after 4 min of continuous application of D600 is labelled 3. In the control there is a large inhibition (a3), whereas in the presence of internal GTP- γ -S, there is still potentiation, although to a lesser extent than at 30 s (b3). The histogram in c represents the mean $\pm$ s.e.m. of the results from the number of cells shown in parentheses for application of D600 for 30 s (left) or 4 min (right). The data have been normalized, the maximum calcium channel current in each cell before D600 application being considered unity. Their actual mean amplitude was 2.8 ± 0.6 nA ($n = 5$) in control cells and 0.60 ± 0.19 nA ($n = 6$) in GTP- γ -S-containing cells. Pre-treatment of cells with pertussis toxin (500 ng ml $^{-1}$ for 2–3 h at 37°C before the start of the experiment) prevented D600-induced potentiation in the presence of GTP- γ -S. This concentration of pertussis toxin produced complete ADP ribosylation of α_i and α_o in these cells (unpublished results in collaboration with Dr G. Milligan, University of Glasgow). (Pertussis toxin was supplied by Dr L. Irons, PHLS-CAMR Porton Down, UK). (Before D600, $\square$; control, $\blacksquare$; GTP- γ -S, $\boxtimes$; pertussis toxin+GTP- γ -S, $\square$). The current-voltage (I/V) relationships illustrated in d were performed in a cell with internal GTP- γ -S, before application of D600 ($\blacktriangle$) and during D600 application ($\blacksquare$). No shift was observed in the I/V curve in the presence of D600. Currents were activated in 10 mV steps from V_H -80 mV, and the enhancement of the peak current in the presence of D600 is thus an underestimate because of the development of frequency-dependent inhibition. The respective maximum calcium channel currents are shown beneath the graph. **Methods.** DRGs were grown in culture and calcium channel currents were recorded as previously described⁴ using Ba $^{2+}$ as the charge carrier. Patch pipettes of 1–4 M Ω resistance were used and the patch-pipette solution contained *inter alia* 2 mM Mg ATP. In control experiments, activating currents at 0.03 Hz, rundown of calcium channel currents was minimal over 20 min. When GTP- γ -S was included in the patch pipette the initial current was recorded immediately after establishing the whole-cell clamp, and the cell was then equilibrated for 5 min before beginning the experiment. Drugs were applied by low pressure and currents were leak-subtracted as previously described⁴.

still increased in all cells (by $58 \pm 31\%$; $n = 8$). Whereas the inhibition of control calcium channel currents by nifedipine was not accompanied by a shift in the voltage dependence of current activation ($n = 4$), the enhancement in the presence of internal GTP- γ -S was often associated with a 5–10 mV hyperpolarizing shift (7/10 cells). This is similar to the stimulatory response observed to Bay K 8644 (refs 9, 14), suggesting that the two

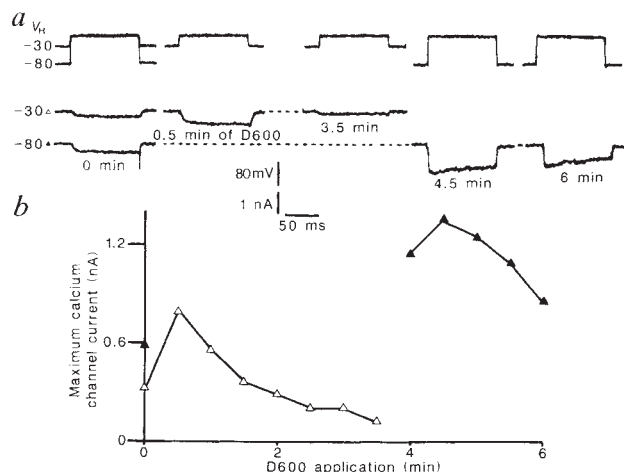


Fig. 2 The voltage dependence of the response to D600 in the presence of internal GTP- γ -S. Data is shown from a cell recorded with GTP- γ -S (500 μ M) in the patch pipette. In a the upper traces show the voltage step commands, and the lower traces the respective maximum calcium channel currents. On the left are the currents evoked from V_H -80 and -30 mV before application of D600. The cell was then held at -30 mV while 10 μ M D600 was continuously applied and currents were evoked at 0.03 Hz. The numbers under the current traces indicate the time after the start of drug application. The complete time course of the experiment is shown in b where the amplitude of the maximum calcium channel current is plotted against time after start of application of D600 at V_H -30 mV ($\triangle$) or V_H -80 mV ($\blacktriangle$). D600 produced a transient enhancement (second trace in a) followed by a use-dependent inhibition of the calcium channel current (third trace in a). Between 3.5 and 4.0 min, V_H was changed from -30 mV to -80 mV while application of D600 was continued. The calcium channel current was immediately potentiated by D600 (fourth trace in a) compared to the control current evoked from V_H -80 mV before the start of drug application. This potentiation gradually declined (fifth trace in a; $\blacktriangle$ in b).

dihydropyridines enhance calcium channel currents by similar mechanisms.

Separate binding sites have been identified for the different calcium channel ligands⁶, and there are also several sites of differing affinity for the dihydropyridines¹⁵. It has been suggested that their antagonist action on calcium currents in heart muscle might be mediated by a high-affinity site, and that their agonist action might be due to the interconversion of the same site¹⁶. The effects of Bay K 8644 are mediated in part by a low-affinity site¹⁶. Previously, binding studies have also indicated the possibility of an interaction between G proteins and the binding sites for the calcium channel ligands^{9,17}.

It is unlikely that the present response to GTP- γ -S is due to non-specific events unrelated to G-protein activation, as similar effects on calcium channel currents are produced by a much lower concentration of GTP- γ -S (20 μ M) liberated from a precursor by flash photolysis¹⁸. We have observed similar stimulatory responses to D600 with a 10-fold lower concentration of GTP- γ -S in the patch pipette (50 μ M; $n = 4$), and the effect of GTP- γ -S is mimicked by another GTP analogue, guanylylimidodiphosphate, and is blocked by pertussis toxin. It is not observed when the GDP analogue GDP- β -S or the ATP analogue ATP- γ -S is substituted for GTP- γ -S. The ability of pertussis toxin to prevent the response to GTP- γ -S requires some comment. ADP ribosylation of G_i/G_o by pertussis toxin does not prevent dissociation of ADP ribosylated α from $\beta\gamma$ ¹⁹, but it does slow the rate of activation of G_i/G_o by guanine nucleotide analogues²⁰, and in some cell types it has been found to prevent the responses to these analogues, as well as abolishing the effect of GTP and hormones²¹.

The actions of calcium channel antagonists, particularly the

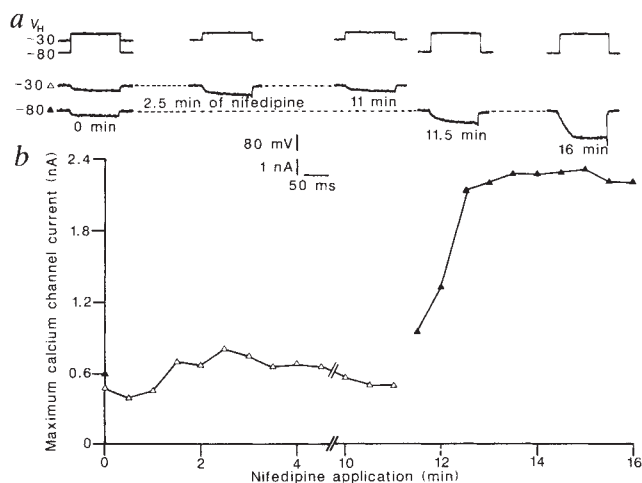


Fig. 3 The voltage dependence of the response to nifedipine in the presence of internal GTP- γ -S. The experimental design is similar to that of Fig. 2. In *a*, the left-hand current traces show the maximum calcium channel currents evoked at $V_H = -80$ mV and -30 mV before application of nifedipine. The cell was held at -30 mV and 5 μ M nifedipine was then applied continuously. It produced a transient potentiation (second and third current traces in *a*, Δ in *b*). Between 11 and 11.5 min, V_H was changed to -80 mV and an immediate potentiation was observed which increased over 2 min (fourth and fifth traces in *a*, $\blacktriangle$ in *b*). This potentiation was accompanied by a 5 mV hyperpolarizing shift for activation of the maximum calcium channel current.

dihydropyridines, are strongly dependent on membrane potential, and it has been suggested that changes in potential may induce interconversion between the agonist and antagonist forms of the high-affinity binding site¹⁶. In these experiments, the D600-induced potentiation of the calcium channel current in the presence of GTP- γ -S, recorded at $V_H = -80$ mV, could rapidly be converted to an inhibition when V_H was depolarized to -30 mV ($n = 5$). The reverse experiment is shown in Fig. 2. Before D600 application, the calcium channel current with GTP- γ -S was similar at $V_H = -30$ mV and $V_H = -80$ mV (Fig. 2*a*, first set of current traces). The cell was then held at -30 mV, and D600 produced an initial transient increase in the calcium channel current, to $233 \pm 24\%$ of its control level ($n = 5$). After 3.5 min of D600 application at $V_H = -30$ mV the calcium channel current was then inhibited by $69 \pm 17\%$ ($n = 5$) (Fig. 2*a*, second and third traces; Fig. 2*b*, open triangles). Between 3.5 and 4 min after the start of D600 application, the holding potential was changed to $V_H = -80$ mV, and 1 min later the current, recorded at the same clamp potential $V_c(0$ mV), was potentiated by $256 \pm 35\%$ ($n = 5$) compared to that at $V_H = -80$ mV before D600 application (Fig. 2*a*, fourth trace; Fig. 2*b* closed triangles). The gradual reduction of the calcium channel current following its potentiation by D600 (Fig. 2*a*, fifth trace) is probably due to the frequency dependence of the inhibition by D600 (ref. 8), and stabilization of the inactivated channel state²².

When this experiment was repeated with nifedipine, similar results were obtained, although the effects were slower in onset and less inhibition was observed (Fig. 3). At $V_H = -30$ mV, nifedipine produced a small potentiation (Fig. 3*a*, second current trace, compared to left-hand trace before nifedipine application; and Fig. 3*b*, open triangles). The maximum potentiation at $V_H = -30$ mV was 56% (from 0.66 ± 0.21 nA to 1.03 ± 0.16 nA, $n = 5$), and slowly declined to control (3/5 cells) or below control levels (2/5 cells). When V_H was then changed to -80 mV (between 11 and 11.5 min in Fig. 3), an immediate potentiation was observed which increased over the next 2 min and then remained fairly constant (Fig. 3*a*, fourth and fifth traces; Fig. 3*b*, closed triangles). The maximum potentiation observed at $V_H = -80$ mV, compared to the control amplitude before nifedipine application was 124% (from 1.0 ± 0.4 nA to 2.2 ± 0.6 nA, $n = 3$). It is unlikely that this potentiation results from the calcium channel ligand reversing the inhibitory effects of activated G protein on N currents. Nifedipine is not thought to interact with N channels¹³, and the potentiated current, unlike N currents⁵, is non-inactivating. In several experiments the sustained current potentiated by either calcium channel ligand was larger than that observed initially before inhibition by GTP- γ -S.

In a previous study, Brown *et al.*¹⁶ suggested that inhibitory responses to dihydropyridine antagonists such as nitrendipine are produced by their binding to a high-affinity site on the inactivated channel favoured at depolarized holding potentials. Stimulatory responses are produced by the same ligands binding to the same site on the (presumably conformationally altered) channel in its resting state, favoured by hyperpolarized potentials. In the resting state the channel will be available to open upon depolarization, whereas in the inactivated state it will not. The present finding that the stimulatory response to several calcium channel antagonists is massively enhanced by internal GTP- γ -S could be accounted for by the hypothesis that activated G_i/G_o associates with and stabilizes the resting state of the channel, acting in concert with membrane hyperpolarization. This would also explain why the stimulatory response in the presence of GTP- γ -S is reduced by depolarization as at these holding potentials there is a greater steady-state inactivation of the channels. That the stimulatory response requires an activated G protein as well as hyperpolarized membrane potentials is indicated by the prevention of the agonist response by pertussis toxin. In the absence of a non-hydrolysable GTP analogue, a requirement for endogenous GTP might explain the failure of some previous studies to observe any agonist effects in response to the calcium channel antagonist ligands⁸, and the transient and variable potentiation observed by others^{9,16,23}. It may also explain some of the differences in response to calcium channel antagonists between tissues²⁴. In several systems, the inhibition of calcium currents by internal GTP- γ -S mimics the effect of a receptor agonist^{2,4,25}. We are now examining whether these receptor agonists modify the response to calcium channel ligands and whether this response results from a direct interaction with the calcium ion channel itself.

This work was supported by the MRC.

Received 11 September; accepted 23 November 1987.

- Holz, G. G., Rane, S. G. & Dunlap, K. *Nature* **319**, 670-672 (1986).
- Scott, R. H. & Dolphin, A. C. *Neurosci. Lett.* **56**, 59-64 (1986).
- Hescheler, J., Rosenthal, W., Trautwein, W. & Schultz, G. *Nature* **325**, 445-447 (1987).
- Dolphin, A. C. & Scott, R. H. *J. Physiol., Lond.* **386**, 1-17 (1987).
- Nowicky, M. C., Fox, A. P. & Tsien, R. W. *Nature* **316**, 440-443 (1985).
- Striessnig, J., Goll, A., Moosburger, K. & Glossman, H. *FEBS Lett.* **197**, 204-210 (1986).
- Triggler, D. J., Skattebol, A., Rampe, D., Joslyn, A. & Gengo, P. in *New Insights into cell and membrane transport processes*. (eds Poste, G. & Crooke, S. T.) 125-142 (Plenum, New York, 1986).
- Lee, K. S. & Tsien, R. W. *Nature* **302**, 790-794 (1983).
- Hess, P., Lansman, J. B. & Tsien, R. W. *Nature* **311**, 538-544 (1984).
- Dolphin, A. C. *Trends Neurosci.* **10**, 53-57 (1987).
- Nowicky, M. C., Fox, A. P. & Tsien, R. W. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2178-2184 (1985).
- Brown, A. M., Kunze, D. L. & Yatani, A. *Nature* **311**, 570-572 (1984).
- Fox, A. P., Nowicky, M. C. & Tsien, R. W. *J. Physiol., Lond.* **394**, 149-172 (1987).
- Dolphin, A. C. & Scott, R. H. *Br. J. Pharmac.* **90**, 14P (1987).
- Rogart, R. B., DeBruyn Kops, A. & Dzau, V. J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7452-7456 (1986).
- Brown, A. M., Kunze, D. L. & Yatani, A. *J. Physiol., Lond.* **379**, 495-514 (1986).
- Galizzi, J.-P., Fosset, M. & Lazdunski, M. *Eur. J. Biochem.* **144**, 211-215 (1984).
- Dolphin, A. C., Scott, R. H., Trentham, D. R. & Wootton, J. F. *J. Physiol., Lond.* **390**, 89P (1987).
- Huff, R. M. & Neer, E. J. *J. biol. Chem.* **261**, 1105-1110 (1986).
- Jakobs, K. H., Aktories, K. & Schultz, G. *Eur. J. Biochem.* **140**, 177-181 (1984).
- Murayama, T., Katada, T. & Ui, M. *Archs Biochem. Biophys.* **221**, 381-390 (1983).
- Bean, B. P. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6388-6392 (1984).
- Schwartz, A., Grupp, I., Grupp, G., Williams, J. & Vaghy, P. *Biochem. biophys. Res. Commun.* **125**, 387-394 (1984).
- Miller, R. J. *Science* **235**, 46-52 (1987).
- Lewis, D. L., Weight, F. F. & Luini, A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9035-9039 (1986).

HLA-A2 peptides can regulate cytotoxicity by human allogeneic T lymphocytes

Carol Clayberger*, Peter Parham†, Jonathan Rothbard‡, David S. Ludwig§, Gary K. Schoolnik§ & Alan M. Krensky*||

* Departments of Pediatrics, † Cell Biology and § Medical Microbiology, Stanford University School of Medicine, Stanford, California 94305, USA

‡ Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, UK

The class-I and class-II molecules encoded by the major histocompatibility complex (MHC) are homologous proteins which allow cytotoxic and helper T cells to recognize foreign antigens^{1,2}. Recent studies have shown that the form of the antigen recognized by T cells is generally not a native protein but rather a short peptide fragment^{3,4} and that class-II molecules specifically bind antigenic peptides^{5,6}. Furthermore, the three-dimensional structure of the human MHC class-I molecule, HLA-A2, is consistent with a peptide-binding function for MHC class-I molecules^{7,8}. An outstanding question concerns the molecular nature and involvement of MHC-bound peptides in antigens recognized by alloreactive T cells. In this study the effects of peptides derived from HLA-A2 on cytotoxicity of alloreactive cytotoxic T cells (T_C) cells are presented. Peptides can inhibit lysis by binding to the T cell or sensitizing to lysis by binding an HLA-A2-related class-I molecule (HLA-Aw69) on the target cell. Thus, allospecific T_C cells can recognize HLA-derived peptides in the context of the MHC.

Previously we showed that peptides corresponding to the region around tryptophan 107 in the α_2 domain of HLA-A2 (Fig. 1) specifically inhibit lysis of many HLA-A2-specific T_C cells (ref. 9). Subsequent analysis revealed that an octapeptide corresponding to residues 101–108 was sufficient to give good inhibition (data not shown). Tryptophan 107 is on a bend, away from the peptide-binding region and is critical for an important serological epitope of HLA-A2 (refs 10 and 11). A second important epitope involves residues 62–65 of the α_1 -helical region of the α_1 domain^{7,8}. This epitope, originally defined with the monoclonal antibody (MAb) MA2.1 (ref. 12), is shared by all subtypes of HLA-A2 and HLA-B17 (ref. 13). Peptide A2.56–69 derived from this region did not affect lysis by HLA-A2 specific T_C cells⁹. We therefore asked whether T_C cells with specificity for HLA-A2 and HLA-B17 could be obtained and if their cytotoxicity was inhibited by this peptide. T_C-cell clones specific for the epitope shared by HLA-A2 and HLA-B17 were derived^{14,15} and one, designated clone A2/B17, was studied in detail. Another clone specific for HLA-B17 but not HLA-A2 was designated clone B17. The target specificity of clone A2/B17 and the finding that cytotoxicity was blocked by MAb MA2.1 indicated that this clone recognized the epitope shared by HLA-A2 and HLA-B17.

The effects of various peptides on cytotoxicity by these clones and an HLA-A2-specific T_C-cell clone is shown in Fig. 2. Peptide A2.56–69, which encompasses the shared serological epitope¹³, specifically inhibited the killing of both HLA-A2 and HLA-B17-expressing target cells by clone A2/B17. In contrast, this peptide had no effect upon the lysis of HLA-B17-expressing cells by clone B17. Clone A2/B17 was not inhibited by a peptide derived from residues 56–69 of HLA-Aw68.1 or a series of unrelated peptides. The A2.98–113 peptide did not affect the lysis of HLA-B17-expressing targets by clone A2/B17, but some inhibition was observed at high concentrations with HLA-A2-expressing targets. This difference indicates that the epitopes of HLA-A2 and HLA-B17 recognized by clone A2/B17 are not

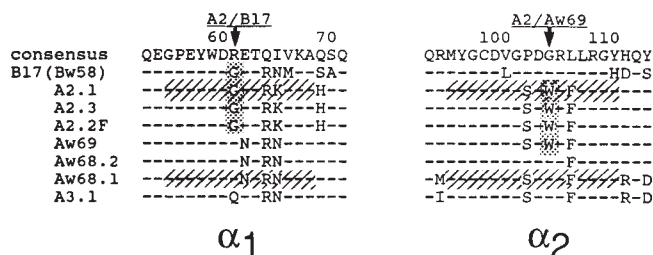


Fig. 1 The epitopes studied and peptides used in these experiments. Selected protein sequences from the α_1 and α_2 domains of eight HLA-A, B molecules are shown using the standard one-letter amino-acid code. The sequence of the HLA-Bw58 subtype of HLA-B17 is from Ways *et al.*²³, that of HLA-A3.1 is from Strachen *et al.*²⁴ and the remaining sequences of the HLA-A2/28 family are from Holmes *et al.*²⁵. Peptides A2.56–69 and Aw68.56–69 from α_1 , and A2.98–113 and Aw68.98–113 from α_2 are indicated by cross-hatching. The two residues found to be critical for the epitopes shared by subtypes of HLA-A2 and HLA-B17 (glycine 62) and subtypes of HLA-A2 and HLA-Aw69 (tryptophan 107) are indicated by stipple and the vertical arrows. The consensus sequence is derived from a total of 23 HLA-A, B and C sequences.

precisely the same. As previously reported, the HLA-A2-specific T_C cells were inhibited only by the A2.98–113 peptide⁹. These results show that (1) the capacity of peptides to inhibit alloreactive T_C cells is not restricted to the region involving residue 107 of the α_2 domain and (2) the sequence of the inhibitory peptide correlates with the epitope specificity of the T_C cells.

Do these peptides inhibit lysis by interacting with the T_C cells, the target cells, or with both? To answer the question, T_C cells and/or target cells were pretreated with either A2.98–113 or Aw68.98–113 for 30 min, washed and used in a ⁵¹Cr-release assay (Table 1, panel a). Lysis was inhibited when the T_C cells, but not the target cells, were pretreated with A2.98–113. No inhibitory effects were observed when T_C cells or target cells were pretreated with the control-peptide Aw68.98–113. Similar results were found for the A2.56–69 peptide (data not shown).

As it has been shown that viral^{16,17} and xenogeneic¹⁸ peptides can be recognized by T_C cells in the context of MHC class-I antigens, it was of interest to establish whether allospecific T_C cells recognize intact HLA molecules or peptides derived from HLAs. In initial studies, we added the A2.98–113 peptide to B-cell lines which do not express HLA-A2 and asked whether they could be recognized and lysed by HLA-A2-specific T_C cells.

Table 1 Locus of peptide effect

a, Inhibition of cytotoxicity		
Pretreatment		Specific cytotoxicity (%)
T _C cell-A2	Target*	
Medium	Medium	55
A2.98–113	Medium	28
Medium	A2.98–113	52
A2.98–113	A2.98–113	24
b, Sensitization to cytotoxicity		
Pretreatment		Specific cytotoxicity (%)
Clone A2/B17	Target†	
Medium	Medium	10
A2.56–69	Medium	15
Medium	A2.56–69	47
A2.56–69	A2.56–69	24

10⁶ T_C cells or ⁵¹Cr-labelled targets were treated for 30 min with medium or 150 μ M peptide at 37°C, washed three times and used in ⁵¹Cr-cytotoxicity assay.

* Target: JY(HLA-A2, 2: B7, 7, DR4, 6), E:T = 1:1.

† Target: IDF (HLA-Aw69, 26; B18, 38; DR5), E:T = 5:1.

|| To whom correspondence should be addressed.

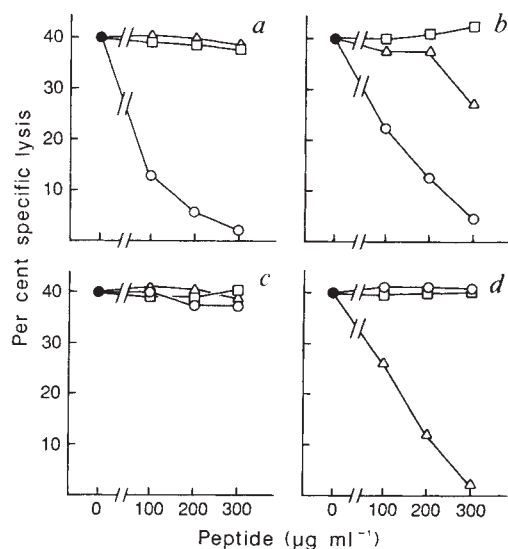


Fig. 2 Peptides from the α_1 or α_2 domains of HLA-A2 can specifically inhibit cytolysis.

Methods. The T_C cells used for these experiments were clone A2/B17 (panels *a* and *b*); clone B17 (panel *c*) or T_C -cell-A2 (panel *d*). All clones were derived from the peripheral blood lymphocytes of a normal donor (HLA-A3; B7; DR6) essentially as described^{14,15}. For clone A2/B17, the cells were stimulated in primary culture with the irradiated (10,000 Rad) B-lymphoblastoid cell-line Mag (HLA-A26, 33; B17, 51) and cloned using the SB cell-line (HLA-A1, 2; B17, 44; DR2, 6) as a stimulator. Clone B17 was derived from the same primary culture as clone A2/B17 but was cloned using the SH cell line (HLA-A3, w33; B7, 17(w57)) as a stimulator. T_C -cell-A2 was derived from cells stimulated in primary culture with the JY cell line and cloned using the Herluff cell line (HLA-A2; B12, 35; DR4, 7) as a stimulator. T_C cells were incubated with the shown concentrations of peptides (○), A2.56-69; (□), Aw68.56-69; (△), A2.98-113, or (■), Aw68.98-113 for 20 mins before the addition of 10^3 ^{51}Cr -labelled T7529 cells (HLA-Aw33; B17(w58); DR6) or JY cells (HLA-A2; B7; DR4, 6). The final molar concentrations of peptides used in the assay at 100 μg per ml were 4.9×10^{-5} M for A2.98-113; 5.2×10^{-5} M for Aw68.98-113; 5.9×10^{-5} M for A2.56-69; and 5.9×10^{-5} M for Aw68.56-69. The peptides were present throughout the cytotoxicity assay which was performed as described^{19,20}. Peptides were prepared as stock solutions at 1 mg per ml in phosphate-buffered saline and diluted in complete medium (minimum essential medium supplemented with 10% calf serum) to give the final concentration. In all cases, effector-to-target ratios were 1:1. Clone A2/B17 is different from the PWSB line previously described⁹. The PWSB line, which lysed both HLA-A2 and HLA-B17-expressing target cells, was polyclonal and was not inhibited by the A2.56-69 peptide. This T_C -cell line probably was a mixture of T_C cells including individual clones specific for HLA-A2 or HLA-B17.

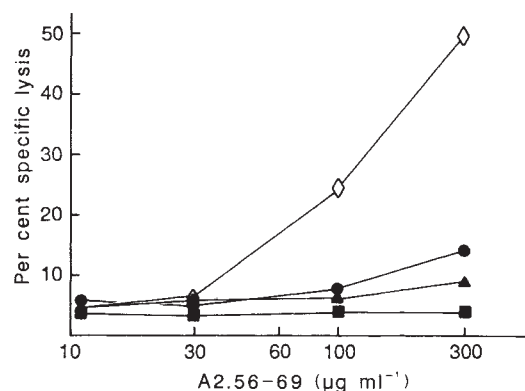


Fig. 3 Sensitization of targets to lysis by peptide A2.56-69. Clone A2/B17 was incubated with the indicated concentration of peptide A2.56-69 and ^{51}Cr -labelled targets at an effector-to-target ratio of 5:1 for 5 h. Targets were (■), IBW4 (HLA-A3; B35, DR1); (▲), LB (HLA-Aw68.1; B40, DR6); (●), Pally (HLA-Aw68.2, 26; B14, 38; DR1, 4), or (◇), IDF (HLA-Aw69, 26; B18, 38, DR5).

the epitope of HLA-A2 and HLA-B17 formed by residues 56-69 but does not bind to HLA-Aw69 or peptide A2.56-69, did not inhibit the lysis. To assess whether sensitization resulted from peptide interaction with the T_C cells or the target cells, cells were pretreated with 150 μM A2.56-69, washed, and then tested for cytolysis (Table 1, panel *b*). Targets expressing HLA-Aw69 were lysed when the targets but not the T_C cells, were pretreated with A2.56-69.

The simplest interpretation of these results is that the HLA-A2/B17 specific T_C cells recognize the A2.56-69 peptide in the context of HLA-Aw69 as a restriction element. This implies that the peptide is binding to the HLA-Aw69 molecule. The data and interpretation are similar to those obtained in the influenza^{16,17} and xenogeneic¹⁸ systems and show that alloreactive T_C cells can recognize class-I-derived peptides in a class-I-restricted way.

Peptide A2.56-69 has been shown to regulate the lysis of target cells by two distinct mechanisms. Inhibition occurs at the level of the T_C cells, presumably through competition for the T-cell receptor and is seen for target cells expressing HLA-A2 and/or HLA-B17. Sensitization occurs on the target cell and specifically involves binding of the peptide to the HLA-Aw69 molecule. When excess peptide is present in the assay involving lysis of HLA-Aw69-expressing targets by clone A2/B17 a level of lysis is seen which represents the combination of these two effects (Table 2, panel *b*).

Cells expressing HLA-Aw69 are not lysed by clone A2/B17 unless the A2.56-69 peptide is present, indicating that the HLA-Aw69 molecule plus peptide A2.56-69 form a composite epitope that is seen by the T_C cells. HLA-Aw69 is identical to HLA-A2.1 in the α_2 and α_3 domains. Six amino-acid substitutions are found in the α_1 domain of which five are in the α -helical region (residues 50-84) (refs 7 and 8), three being present in the A2.56-69 peptide (Fig. 1). The crystallographic structure indicates that the peptide-binding site is a deep groove between the α -helical regions of the α_1 and α_2 domains^{7,8}. If peptide A2.56-69 were to adopt an oriented α -helical structure in the binding site of HLA-Aw69 then α helices derived from the α_1 and α_2 domains of HLA-A2.1 would be juxtaposed. It is in this way that HLA-Aw69 plus peptide may mimic the native structure of HLA-A2.1.

Residues 56-69 lie in an α helix which forms part of the peptide-binding site, whereas residue 107 is part of a turn between two strands of β structure at some distance from the α helices and peptide-binding region^{7,8}. The A2.98-113 peptide may maintain elements of this structure in solution and have

No sensitization to lysis could be detected using target cells expressing a variety of HLA molecules and a wide range of peptide concentrations (0.005-150 μM) (data not shown). We also tested whether peptides derived from HLA-A2 could sensitize target cells expressing molecules closely related to HLA-A2 such as HLA-Aw68.1, HLA-Aw68.2 and HLA-Aw69 (Fig. 1). Clone A2/B17 does not lyse target cells expressing HLA-Aw68.1, HLA-Aw68.2 or HLA-Aw69 as the critical residues around position 62 differ from those found in HLA-A2 and HLA-B17. But when peptide A2.56-69 was included in the cytotoxicity assay, significant lysis of HLA-Aw69-expressing targets by clone A2/B17 was seen (Fig. 3). In contrast, targets expressing HLA-Aw68.1, HLA-Aw68.2 or the unrelated HLA-A3 molecule were not lysed. Lysis was blocked by MAb CR11-351, which binds only to the HLA-Aw69 molecule of the target cell, directly showing the involvement of HLA-Aw69 in the sensitization by peptide A2.56-69. In contrast, the MAb MA2.1, which binds to

little affinity for the peptide-binding site of class-I molecules. This interpretation could explain why peptides corresponding to this region are inhibitors of HLA-A2 directed cytolysis but cannot induce cytolysis.

What is the *in vivo* relevance of these findings? Cells express up to six class-I HLA molecules providing ample opportunity for peptides derived from one molecule to bind to another. Peptides could be derived from endogenous HLA molecules as products of the normal pathways of biosynthesis and intracellular degradation^{20,21}. Alternatively, they could result from endocytosis of the soluble HLA molecules found in serum²². The number of T-cell epitopes which could be generated by HLA peptides associating with intact HLA molecules is potentially large and may in part account for the intensity of the allorresponse.

This work was supported by grants from the American Cancer Society, Weingart Foundation and the NIH. P.P. is a scholar of the Leukemia Society of America; A.M.K. is the recipient of an American Heart Association Clinician-Scientist Award. We thank K. Drickamer for amino-acid sequence analysis, Pamela Bjorkman and Don Wiley for sharing their unpublished results, and Linda Hopper for preparation of the manuscript.

Received 7 September; accepted 9 November 1987.

- Schwartz, R. H. *A. Rev. Immun.* **3**, 237-255 (1985).
- Zinkernagel, R. M. & Doherty, P. C. *Adv. Immun.* **27**, 51-177 (1979).
- Ziegler, K. & Unanue, E. R. *J. Immun.* **127**, 1869-1874 (1982).
- Townsend, A. R. M., Gotch, F. M. & Davey, J. *Cell* **42**, 457-468 (1985).
- Buss, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 3968-3971 (1986).
- Babbitt, B., Allen, P., Matsueda, G., Haber, E. & Unanue, E. *Nature* **317**, 359-361 (1985).
- Bjorkman, P. J. *et al. Nature* **329**, 506-511 (1987).
- Bjorkman, P. J. *Nature* **329**, 512-518 (1987).
- Parham, P. *et al. Nature* **325**, 625-628 (1987).
- Salter, R. D., Clayberger, C., Lomen, C. E., Krensky, A. M. & Parham, P. *J. exp. Med.* **166**, 283-288 (1987).
- Layet, C. *et al. J. Immun.* **138**, 2197-2202 (1987).
- McMichael, A. J., Parham, P., Rust, N. & Brodsky, F. *Hum. Immun.* **1**, 121-138 (1980).
- Ways, J. P. & Parham, P. *Biochem. J.* **216**, 423-432 (1983).
- Clayberger, C. *et al. J. exp. Med.* **162**, 1709-1714 (1984).
- Seong, R., Clayberger, C., Krensky, A. M. & Parnes, J. *J. exp. Med.* (in the press).
- Bastin, J., Rothbard, J. B., Davey, J., Jones, I. & Townsend, A. J. *J. exp. Med.* **165**, 1508-1523 (1987).
- Gotch, F., Rothbard, J., Howland, K., Townsend, A. & McMichael, A. *Nature* **326**, 881-883 (1987).
- Maryanski, J. L., Pala, P., Corradin, G., Jordan, B. R. & Cerottini, J. C. *Nature* **324**, 578-581 (1986).
- Holmes, N. & Parham, P. *EMBO J.* **4**, 2849-2854 (1985).
- Krangel, M. S., Orr, H. T. & Strominger, J. L. *Cell* **18**, 979-991 (1979).
- Ploegh, H. L., Orr, H. T. & Strominger, J. L. *J. Immun.* **126**, 270-275 (1982).
- Gussow, D. & Ploegh, H. *Immunol. Today* **8**, 220-222 (1987).
- Ways, J. P., Coppin, H. & Parham, P. *J. Biol. Chem.* **260**, 11924-11930 (1985).
- Strachen, T., Sodoyer, R., Damotte, M. & Jordan, B. R. *EMBO J.* **3**, 887-896 (1984).
- Holmes, N., Ennis, P., Wan, A., Denney, D. & Parham, P. *J. Immun.* **139**, 936-941 (1987).
- Krensky, A. M., Reiss, C. S., Mier, J. W., Strominger, J. L. & Burakoff, S. J. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2365-2370 (1982).

Importance of FSH-releasing protein and inhibin in erythrodifferentiation

John Yu*, Li-en Shao*, Victor Lemas*, Alice L. Yu†, Joan Vaughan‡, Jean Rivier‡ & Wylie Vale‡

* Department of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, California 92037, USA

† Department of Pediatrics, University of California at San Diego, La Jolla, California 92093, USA

‡ Clayton Foundation Laboratories for Peptide Biology, The Salk Institute for Biological Studies, La Jolla, California 92037, USA

Inhibin is a hypophysiotropic hormone which selectively suppresses the secretion of pituitary follicle-stimulating hormone. It has been isolated from gonadal fluids and characterized as a protein heterodimer consisting of an α subunit and one of two β subunits (β_A or β_B)¹⁻⁴. FSH-releasing protein (FRP), also named activin, is a dimer consisting of two inhibin β -chains⁵⁻⁶. A factor from conditioned medium of a leukaemia cell line has been isolated which can induce mouse Friend cells to become benzidine-positive, and which shares a similar N-terminal sequence with porcine FRP⁷. In this report, we find that FRP and inhibin modulate both the induction of haemoglobin accumulation in a human erythroleukaemic cell line, K562, and the proliferation of erythroid progenitor cells in human bone marrow culture. These two proteins could constitute a novel humoral regulatory control of erythropoiesis which would involve two types of related protein dimers with functionally opposite effects.

K562 cells provide a model system for studying human erythroid differentiation as they can be induced by haemin or other chemicals to undergo differentiation and accumulate haemoglobins⁸⁻¹¹. We have purified inhibin from ovine rete testis fluid by immunoaffinity chromatography (J.V. *et al.* manuscript in preparation) and HPLC to homogeneity, and have identified it as an $\alpha\beta_A$ heterodimer³; FRP (activin A) was purified from porcine follicular fluid and characterized as a $\beta_A\beta_A$ homodimer⁵. Figure 1 shows that incubation of K562 cells with FRP for three days did not significantly affect proliferation and viability of the cells. A clonal assay for K562 cells was performed in 0.8% methylcellulose¹² containing FRP at 1, 5 and 10 ng ml⁻¹: after five days of culture the number of colonies in these cultures

Table 1 Haemoglobin contents in control and induced K562 cells

Inducers	Haemoglobin content/10 ⁶ cells μ g	Benzidine-staining cells(%)
None	0.06	6
FRP, 10 ng ml ⁻¹	1.37	48
Haemin, 25 μ M	1.50	94

Cell lysates were prepared from control, FRP- or haemin-induced K562 cells; $\sim 1 \times 10^7$ cells were washed twice in PBS, pH 7.4, and aliquots taken for benzidine staining¹⁸. One vol packed cells was then mixed with 1 vol deionized water and 0.5 vol CCl₄, centrifuged at 4°C, and the amount of haemoglobin in the supernatants was determined as described¹³.

was respectively 97, 90 and 92% of the number for a control culture without FRP, showing that FRP has no cytotoxic effect on K562 cells. The FRP cultures, however, contained many colonies that were smaller in size (≤ 16 cells per colony) than those in the control culture. The cells in the smaller colonies were positive for haemoglobin staining with benzidine, and the large colonies were negative, except for a few of the mixed colonies. These results indicate that prolonged exposure of K562 cells to FRP causes them to become terminally differentiated and limits their proliferation.

Also shown in Fig. 1 is the response of K562 cells after incubation with FRP in liquid culture for three days, when they become benzidine-positive in a dose-responsive manner: maximum induction was observed with FRP at 10 ng ml⁻¹ (≈ 300 pM), with $\sim 56\%$ cells becoming benzidine-positive; higher concentrations of FRP did not further increase the percentage of cells induced. The effective dose for half-maximum induction by FRP was ~ 1 ng ml⁻¹ (≈ 30 pM), indicating that this protein was highly potent. Haemoglobin content of lysates from control, FRP- or haemin-induced K562 cells⁹ were quantified by spectrophotometric assay¹³. Cells treated with FRP had ~ 23 -fold increase in haemoglobin content (Table 1). They produced nearly as much haemoglobin as the haemin-induced sample, but contained only 48% benzidine-staining cells. Therefore, for the responsive K562 cells, FRP is about twice as effective as haemin in inducing the production of haemoglobin on a single cell basis. The presence of haemoglobin was confirmed and semi-quantitative estimates of its accumulation made by

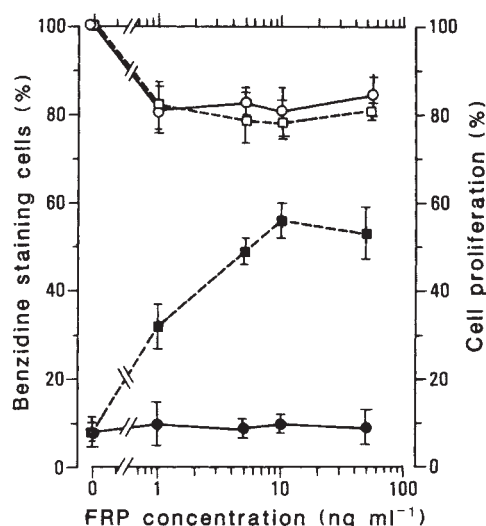


Fig. 1 Effect of FRP and buffer on proliferation and differentiation of K562 cells. K562 cells were incubated with FRP to the final concentration specified for three days, aliquots were then analysed for cell density ($\square$ — $\square$) and percentage of cells staining for haemoglobin with benzidine ($\blacksquare$ — $\blacksquare$). Equal amounts of HPLC buffer were added to control cultures, and data from these cultures on cell density and percentage of haemoglobin-containing cells are represented as $\circ$ — $\circ$ and $\bullet$ — $\bullet$ respectively.

Methods. Stock cultures of K562 were grown in RPMI 1640 medium supplemented with 50 IU penicillin per ml, 50 μ g streptomycin per ml, and 15% fetal calf serum²⁰. Cell cultures at density of 1.7×10^5 per ml were incubated with FRP from a stock solution of $1 \mu\text{g ml}^{-1}$ in 0.1% bovine serum albumin, 2.8% acetonitrile, 0.004% trifluoroacetic acid. Cell numbers were counted with a haemocytometer. The effect on cell proliferation was presented as percentage of cell number after incubation with different doses of FRP, using the value (7.1×10^5 per ml) for the control sample without FRP as 100%. Cell viability was determined by Trypan blue dye exclusion and gave more than 98% viable cells. The number of cells containing haemoglobin was assayed by benzidine staining¹⁸. All data were presented as mean value $\pm$ s.d.

polyacrylamide gel electrophoresis, followed by immunochemical staining of equal amounts of lysates prepared from FRP- and haemin-treated cells with antibodies directed against human haemoglobins (results not shown).

We have previously reported that purified porcine inhibin lowers baseline production of FSH, and thereby obscures the effects of the lower concentrations of FRP⁵. In Table 2, we analysed the effect of inhibin upon induction of differentiation in K562 cells in the presence or absence of FRP, based on benzidine staining for haemoglobin. It was found that addition of inhibin suppressed the induction of differentiation by FRP in a dose-related manner, with a 50% effective inhibitory concentration in the range 1 – 5 ng ml^{-1} . Inhibin alone seems to decrease the spontaneous differentiation of K562 cells in culture. Statistical analysis of variance indicated that induction of erythroid differentiation was significantly stimulated by FRP and inhibited by inhibin. It also showed a functional interaction between the two proteins. These conclusions were confirmed by immunocytochemical analyses of haemoglobin accumulation in K562 cells after treatment with FRP or inhibin or both (manuscript in preparation).

In human bone marrow cultures, proliferation and differentiation of erythroid progenitors were analysed by scoring haemoglobinized colonies with unique morphology characteristics of erythroid colony forming unit (CFU-E) after 7 days of culture (Table 3). The addition of FRP significantly potentiated the formation of CFU-E in the presence of erythropoietin up to 300%. FRP and/or inhibin alone did not promote colony forma-

Table 2 Effect of FRP and inhibin on induction of K562 cells to differentiation

Inhibin concentration (ng ml ⁻¹)	FRG concentration (ng ml ⁻¹)			
	0	1	5	10
	Benzidine-staining cells (%)			
0	6 $\pm$ 1	23 $\pm$ 3	32 $\pm$ 6	41 $\pm$ 3
1	1 $\pm$ 1	8 $\pm$ 2	21 $\pm$ 2	26 $\pm$ 5
5	3 $\pm$ 2	8 $\pm$ 1	12 $\pm$ 3	16 $\pm$ 2
10	3 $\pm$ 1	3 $\pm$ 2	9 $\pm$ 1	11 $\pm$ 2
50	2 $\pm$ 1	4 $\pm$ 1	6 $\pm$ 4	9 $\pm$ 1

All treatments were added simultaneously and incubation was continued for 72 h. Effect of induction of K562 cells to differentiation was analysed by benzidine staining for haemoglobin¹⁸. Comparisons of data for the effect of inhibin upon FRP-induced accumulation of benzidine-staining cells were made by calculating the critical values of the F -distribution with two-way ANOVA analysis of variance¹⁴. The critical values for effects of FRP and inhibin treatments on induction of differentiation are $F_{[3,40]} = 139.44$ and $F_{[4,40]} = 118.23$, respectively. They are significantly greater than those for 0.1% of the area under F -distribution in ANOVA analysis.

Table 3 Effect of FRP and inhibin on colony formation of CFU-E from human bone marrow

Additives	Units erythropoietin per ml			
	0	0.2	0.5	3.0
	Mean colony number $\pm$ s.d./ 10^5 cells			
FRP, none	0	13.3 $\pm$ 2.5 (100%)	26.3 $\pm$ 2.3 (100%)	41.3 $\pm$ 5.1 (100%)
FRP, 1 ng ml ⁻¹	0	17.7 $\pm$ 2.5 (133%)	34.3 $\pm$ 7.3 (130%)	78.7 $\pm$ 8.5 (191%)
FRP, 5 ng ml ⁻¹	0	24.0 $\pm$ 3.6 (181%)	39.7 $\pm$ 5.6 (151%)	94.3 $\pm$ 7.6 (228%)
FRP, 10 ng ml ⁻¹	0	39.3 $\pm$ 3.5 (296%)	43.6 $\pm$ 6.5 (166%)	121.6 $\pm$ 17.8 (295%)
FRP none +inhibin, 10 ng ml ⁻¹	0	6.3 $\pm$ 2.9 (47%)	18.3 $\pm$ 3.0 (70%)	40.6 $\pm$ 11.0 (98%)
FRP, 1 ng ml ⁻¹ +inhibin, 5 ng ml ⁻¹	0	12.3 $\pm$ 4.0 (93%)	19.0 $\pm$ 3.5 (72%)	54.7 $\pm$ 5.5 (133%)
FRP, 5 ng ml ⁻¹ +inhibin, 10 ng ml ⁻¹	0	9.3 $\pm$ 2.1 (70%)	19.3 $\pm$ 4.0 (73%)	46.7 $\pm$ 3.2 (113%)
FRP, 10 ng ml ⁻¹ +inhibin, 25 ng ml ⁻¹	0	ND	ND	38.3 $\pm$ 6.4 (93%)

Bone marrow culture was carried out with a methylcellulose technique¹⁹ with slight modifications. Briefly, 1 – 2×10^5 mononuclear cells were plated in 35 mm Petri dishes in a 1 ml mixture containing α medium, 0.8% methylcellulose, 30% fetal calf serum, 0.1 mM α -thioglycerol, 50 IU penicillin per ml and 50 μ g streptomycin per ml. The amounts of erythropoietin, FRP and inhibin in each culture are specified. The dishes were incubated at 37 $^{\circ}\text{C}$ in a humidified incubator flushed with 5% CO_2 for 7 days. The dishes were then examined on an inverted microscope for haemoglobinized colonies of CFU-E. Data on the effects of erythropoietin and FRP were subjected to two-way ANOVA statistical analysis¹⁴. The critical values for these two treatments are $F_{[3,32]} = 370.4$, $F_{[3,32]} = 49.5$ respectively. The effect of inhibin was similarly analysed by comparing data from cultures containing erythropoietin and similar amounts of FRP with or without excess inhibin. Numbers within parentheses represent the percentage of CFU-E colony formation relative to control cultures containing similar doses of erythropoietin with no other additives (namely, treatment no. 1); ND, not determined.

tion of CFU-E without erythropoietin. In the absence of FRP, inhibin suppressed the stimulatory effect of low doses (but not high dose) of erythropoietin on CFU-E formation. In the presence of FRP (Table 3, treatments 6–8), the addition of excess amounts of inhibin caused significant reductions of CFU-E colony formation. With FRP at 1 ng ml^{-1} , the reduction was $\sim 30\%$ and became even more prominent at 10 ng ml^{-1} that is, 70% reduction). Therefore, these two protein dimers, FRP and inhibin, may be functionally antagonistic in modulating prolifer-

ation and differentiation of erythroid progenitor cells. At present, it is unknown whether these effects are due to progenitor cells having receptors with overlapping specificities or to separate receptors specific to these two proteins. Alternatively, the suppressive effect of inhibin could be mediated through its interaction with FRP, which could be produced by the cells or be present in the fetal calf serum used in the bone marrow culture. It is also likely that each protein affects erythropoiesis in the absence of the addition of the other, in the same way that they affect FSH secretion in pituitary cell cultures⁵. Regardless of the nature of the interaction, the bone marrow culture experiments indicate that these two hypophysiotropic polypeptides could also provide humoral regulation over erythropoiesis.

The complexity of haematopoietic control is becoming apparent as a result of the availability of highly purified preparations of several haematopoietic growth factors¹⁵. The complete amino-acid sequences of inhibin and FRP have been recently deduced from the complementary DNAs^{16,17}. Here we demonstrate for the first time the regulatory effect of FRP and inhibin on human erythropoiesis *in vitro*; effective concentrations are low (in the picomolar range), indicating that they could also modulate erythropoiesis *in vivo*. Preliminary studies have demonstrated the expression of messenger RNA for the inhibin β_A , but not the α subunit in bone marrow cells of rats (H. Meunier, R. Evans, C. Rivier and W.V., unpublished observations) which suggests that the $\beta_A\beta_A$ dimers could be produced in normal bone marrow to act as a paracrine modulator, whereas inhibin requires an α subunit and may not be formed locally, but could reach the bone marrow from other sources. In fact, rat plasma contains levels of immunoreactive inhibin sufficient to modulate erythropoiesis *in vitro* (C. Rivier, J.V. and W.V., unpublished data). The physiological significance of these proteins and their distributions among different haematopoietic

lineages in human bone marrow remains to be established. Previously, inhibin and FRP were considered to be hypophysiotropic hormones acting on pituitary cells to modulate secretions of pituitary hormones. But our studies, considered with the similarity of sequence and subunit organization to transforming growth factor- β ¹⁶, suggest that inhibin and FRP could have a variety of autocrine, paracrine and hormonal actions in bone marrow, providing a complex humoral regulatory control involving two types of protein dimers with biologically opposite effects in a variety of tissues.

Research in J.Y.'s laboratory was supported by NIH. In addition, research conducted in part by the Clayton Foundation for Research, California Division, and supported in part by NIH, US/AID, and the Population Council.

Received 27 July; accepted 8 October 1987.

- Robertson, D. M. *et al. Biochem. biophys. Res. Commun.* **126**, 220-226 (1985).
- Miyamoto, K. *et al. Biochem. biophys. Res. Commun.* **129**, 396-403 (1985).
- Rivier, J., Spiess, J., McClintock, R., Vaughan, J. & Vale, W. *Biochem. Biophys. Res. Commun.* **133**, 120-127 (1985).
- Ling, N. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 7217-7221 (1985).
- Vale, W. *et al. Nature* **321**, 776-779 (1986).
- Ling, N. *et al. Nature* **321**, 779-782 (1986).
- Eto, Y. *et al. Biochem. biophys. Res. Commun.* **142**, 1095-1103 (1987).
- Lozzio, C. B. & Lozzio, B. B. *Blood* **45**, 321-334 (1975).
- Rutherford, T. R., Clegg, J. B. & Weatherall, D. J. *Nature* **280**, 164-165, (1979).
- Jokinen, M., Gahrberg, C. G. & Andersson, L. C. *Nature* **279**, 604-607 (1979).
- Fukuda, M. *Nature* **285**, 405-407 (1980).
- Rowley, P. T., Betsy, M. O. & Barbara, A. F. *Blood* **65**, 862-868 (1985).
- Tsiftoglou, A. S., Gusella, J. F., Vollock, V. & Housman, D. E. *Cancer Res.* **39**, 3849-3855 (1979).
- Sokal, R. R. & Rohlf, F. J. in *Biometry* 321-371 (Freeman, San Francisco, 1981).
- Metcalf, D. *Science* **229**, 16-22 (1985).
- Mason, A. J. *et al. Nature* **318**, 659-663 (1985).
- Mayo, K. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 5849-5853 (1986).
- Orkin, S. H., Harao, P. I. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **72**, 98-102 (1975).
- Ogawa, M., Parmley, R. T., Bank, H. L. & Spicer, S. S. *Blood* **48**, 407-417 (1976).
- Singh, M. K. & Yu, J. *Nature* **309**, 631-633 (1984).
- Hsu, S. M., Raine, L. & Fanger, H. *Am. J. clin. Pathol.* **75**, 734-738 (1981).

Application of phorbol ester to mouse skin causes a rapid and sustained loss of protein kinase C

Agnes Fournier & Andrew W. Murray

School of Biological Sciences, The Flinders University of South Australia, Bedford Park, South Australia 5042

It is now widely accepted that tumour-promoting phorbol esters activate a Ca^{2+} - and phospholipid-dependent protein kinase (protein kinase C) both *in vitro* and in intact cells¹⁻³, and that the kinase represents a major cellular phorbol ester-binding protein. The phorbol esters act as analogues of diacylglycerol, a natural regulator of protein kinase C, and stabilize the membrane-association of the kinase. Although other molecular targets may exist, protein kinase C activation is probably important in mediating the diverse responses of cultured cells to phorbol esters and in promoting *in vivo* tumours. The enzyme comprises a family of closely related proteins⁴⁻⁷ and has been detected in extracts from mouse epidermal cells⁸⁻¹⁰, the likely targets for two-stage carcinogenesis in mouse skin. In this report we show that application of a single dose of TPA (12-*O*-tetradecanoyl phorbol-13-acetate) to mouse skin results in a rapid and complete loss of protein kinase C activity which is maintained for 3-4 days. This is associated with a loss of immunologically detectable protein kinase C and the accumulation of a smaller protein detectable by antibodies recognizing the regulatory domain of protein kinase C.

Figure 1 shows the effect of a single application of TPA to the dorsal skin of mice on the protein kinase C activity in skin extracts. In the control experiment extracts were prepared 24 hours after the application of acetone; similar results were obtained in thirteen separate experiments. Most of the Ca^{2+} and

phosphatidylserine-dependent protein kinase activity was present in the soluble extract, with a smaller peak of activity in the detergent extract of particulate material. Extracts prepared 24 hours or 72 hours after a single TPA application contained essentially no Ca^{2+} or phospholipid-dependent protein kinase activity. A comparable loss of activity was observed when extracts were prepared 30 minutes after TPA application, whereas after 8 days enzymatic profiles were indistinguishable from the controls (data not shown). Therefore a single application of TPA leads to a rapid loss of protein kinase C activity in mouse skin which is maintained for at least 3 days.

Experiments were also carried out to determine the effects of TPA application on immunologically detectable protein kinase C in mouse skin. The antiserum used in this work (0442) was raised against a synthetic peptide corresponding to amino acids 280-292 of bovine protein kinase C (ref. 11) and has been fully characterized¹². The antibody recognizes the intact protein kinase of relative molecular mass (M_r) ~80,000 (80K) and also interacts with fragments of ~33-35K resulting from limited tryptic proteolysis of protein kinase C. These fragments are believed to contain the Ca^{2+} -, phospholipid- and diacylglycerol/phorbol ester-binding site of the kinase. The other product of limited hydrolysis contains the catalytic domain of protein kinase C and is not recognized by the antibody¹¹. The antibody also cross-reacts with tubulin¹².

Preliminary experiments established that antiserum 0442 reacted with a protein of ~80K present in soluble extracts of mouse skin, and that this protein was not detected in extracts prepared as early as 30 minutes after or up to 3 days after TPA application (data not shown). This result is consistent with the enzymatic data (Fig. 1). Immunologically reactive protein was also present in the particulate fraction of mouse skin (Fig. 2) and 3 major bands of ~80K, 63K and 55K were observed. The reactive proteins of 80K and 55K presumably correspond to intact protein kinase C and tubulin respectively, whereas the

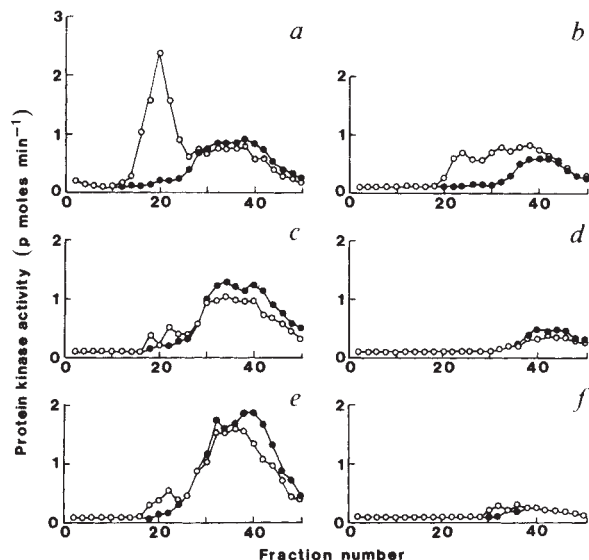


Fig. 1 Effect of TPA application to mouse skin on cytosolic and particulate protein kinase C activity. *a, c, e*, Cytosolic extracts and *b, d, f*, particulate extracts were prepared 24 h after acetone (*a* and *b*), 24 h after TPA (*c* and *d*), or 72 h after TPA (*e* and *f*) application to mouse skin. Extracts were chromatographed on DEAE cellulose and fractions assayed for protein kinase activity in the absence (●) and presence (○) of Ca^{2+} and phosphatidylserine.

Methods. The dorsal skin of female Swiss albino mice (8–16 weeks old) was painted with 20 nmol TPA in 200 μl acetone or with acetone alone. Animals were killed by cervical dislocation and the dorsal skin removed and washed in PBS. The epidermis was scraped off and homogenized in buffer containing 0.25 M sucrose, 20 mM HEPES, 2 mM EDTA, 5 mM EGTA, 2 mM phenylmethylsulphonyl fluoride, 10 mM β -mercaptoethanol and 0.01% leupeptin (pH 7.5; buffer A). The homogenate was filtered through muslin and centrifuged for 60 min at 100,000g. The pellet was resuspended in buffer A containing 0.2% Triton X-100, incubated at 4°C for 60 min and centrifuged as above. Both the cytosolic and the particulate fractions were chromatographed on DE-52 cellulose columns exactly as described before¹⁷ and fractions assayed for protein kinase activity¹⁸. Extracts chromatographed in *a, b, c, d, e* and *f* contained 37.8, 3.5, 59, 8, 58.8 and 5.7 mg of protein respectively. Protein was measured by the method of Lowry¹⁹.

63K protein could be a proteolysis product of protein kinase C (ref. 13). Purified rat brain protein kinase C showed a single major band of about 80K. The 80K protein in mouse skin was undetectable 24 hours after TPA application and this loss was accompanied by the appearance of a series of immunoreactive proteins with M_r between 31K and 35K, a size range consistent with derivation of these fragments from protein kinase C. Although it is possible that they derive from tubulin, there was no obvious loss of 55K material after treatment with TPA.

The experiment shown in Fig. 2 was carried out on isolated particulate material, and possibly there was proteolysis of protein kinase C during sample preparation, despite the presence of chelators and protease inhibitors in the extraction buffer. Therefore skin scrapings were solubilized directly in hot SDS (Fig. 3). Extracts of whole skin contained several proteins which cross-reacted with antibody 0442 as well as the proteins of M_r 80K and 55K. Treatment with TPA resulted in a decrease in 80K immunoreactivity, although the loss was not as rapid or as complete as the apparent loss of enzyme activity. Possibly then, there is some degradation of protein kinase C during the preparation of cellular extracts, or alternatively some of the enzyme is not solubilized from membrane by detergent extraction. But in the experiment in Fig. 3, accumulation of a protein of M_r 31K was observed in samples prepared up to 96 hours after application. It therefore is likely that this fragment is produced in intact cells exposed to TPA and is not an artefact of sample

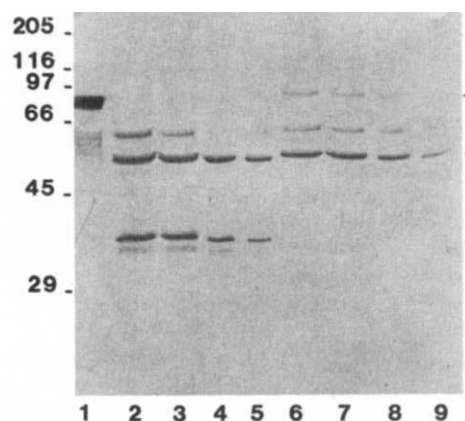


Fig. 2 Immunoblot analysis of particulate skin fractions with antibody against protein kinase C. Lane 1: purified rat brain protein kinase C. Lanes 2–5: particulate fraction prepared 24 h after TPA treatment (550, 396, 198 and 99 μg protein respectively). Lanes 6–9: particulate fraction prepared 24 h after acetone treatment (275, 198, 99 and 49.5 μg protein respectively). M_r of standards are shown on the left (K).

Methods. Particulate fractions were prepared from mouse skin 24 h after application of acetone or 20 nmol TPA (see legend to Fig. 1). SDS polyacrylamide gel electrophoresis (10% gels) was carried out as described²⁰. Proteins were transferred electrophoretically to a nitrocellulose membrane (pore size 0.2 μm) preincubated in PBS containing 5% bovine serum albumin to saturate non-specific binding sites²¹. The primary antibody (0442) to protein kinase C was used at a dilution of 1:1000 and detection was carried out using sheep anti-rabbit antibody coupled to horse radish peroxidase. Controls carried out in parallel using non-immune rabbit serum gave no significant staining. Rat brain protein kinase C was purified essentially as described¹⁶.

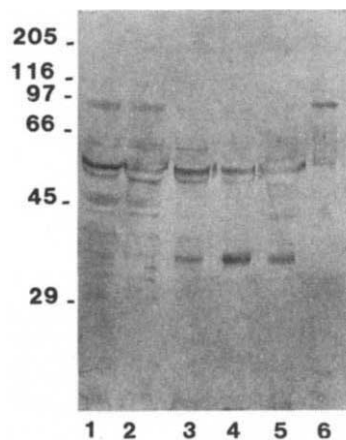


Fig. 3 Immunoblot analysis of hot SDS extracts of whole skin scrapings with antibody against protein kinase C. Lane 1, 24 h after acetone treatment; lane 2, 30 min after TPA treatment; lane 3, 24 h after TPA treatment; lane 4, 72 h after TPA treatment; lane 5, 96 h after TPA treatment; lane 6, purified rat brain protein kinase C. Loadings were all 200 μg protein. M_r of standards are shown on the left. Skin scrapings were extracted directly in hot SDS as described¹³. Other procedures are described in the legend to Fig. 2.

preparation. Formation of the fragment in hot SDS extracts of TPA-treated skin was reproducible and has been observed in three separate experiments. We have no direct evidence that the smaller protein is derived from protein kinase C, but detection of the 80 and 30–35K proteins on immunoblots was abolished by competing protein kinase C (purified from rat brain), whereas staining of the 55K protein was only slightly diminished. Also both the 80 and 30–35K proteins could be detected using a sheep antibody to purified mouse brain protein kinase C (ref.

14) (data not shown). The sheep antibody did not cross-react with tubulin.

We believe that these results could make a major contribution to understanding mouse skin tumour promotion. They should be extended to skin exposed to multiple applications of TPA, and in the standard two-stage carcinogenesis protocol it is usual to apply the phorbol ester twice weekly. The loss of protein kinase C activity and of immunologically reactive protein is probably a result of proteolysis of membrane-associated kinase. This is likely to occur in several cell types exposed to phorbol esters¹⁵ and has been observed in cultured rat skin keratinocytes¹⁶. Evidence was also obtained that proteolysis produces a fragment which contains the Ca^{2+} -, phospholipid- and phorbol ester/diacylglycerol-binding site and which is detectable for several days after a single application of TPA. The possible accumulation of the regulatory domain from protein kinase C is interesting in that the fragment could be biologically active. For example, the amino-terminus of this fragment contains a sequence characteristic of DNA-binding proteins¹¹. As suggested before¹⁵, it is possible that the fragment directly regulates the transcription of specific proteins, and that such regulation is required for the clonal expansion of initiated cells.

We thank Dr Peter J. Parker for providing the antibody 0442, Stephen Hardy for the sample of purified protein kinase C and

Dr Karen Leach for the sheep antibody to protein kinase C. This work was supported by grants from the National Health and Medical Research Council, the Australian Research Grants Scheme and the Universities of South Australia Anti-cancer Foundation.

Received 19 August; accepted 3 November 1987.

1. Nishizuka, Y. *Nature* **308**, 693-697 (1984).
2. Nishizuka, Y. *Science* **233**, 305-312 (1986).
3. Sekar, M. C. & Hokin, L. E. *J. Membrane Biol.* **89**, 193-210 (1986).
4. Coussens, L. *et al. Science* **233**, 859-866 (1986).
5. Knopf, J. *et al. Cell* **46**, 491-502 (1986).
6. Ono, Y. *et al. Science* **236**, 1116-1122 (1987).
7. Housey, G. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 1065-1069 (1987).
8. Garte, S. J., Edinger, F. & Mufson, R. A. *Cancer Lett.* **29**, 215-221 (1985).
9. Gschwendt, M. *et al. Biochem. biophys. Res. Commun.* **124**, 63-68 (1984).
10. Kumar, R. & Holian, O. *J. invest. Dermatol.* **86**, 316-320 (1986).
11. Parker, P. J. *et al. Science* **233**, 853-859 (1986).
12. Stabel, S. *et al. J. cell. Physiol.* **130**, 111-117 (1987).
13. Girard, P. R., Mazzei, G. J. & Kuo, J. F. *J. biol. Chem.* **261**, 370-375 (1986).
14. Cambier, J. C. *et al. Nature* **327**, 629-632 (1987).
15. Murray, A. W., Fournier, A. & Hardy, S. *Trends biochem. Sci.* **13**, 53-54 (1987).
16. Chida, K., Kato, N. & Kuroki, T. *J. biol. Chem.* **261**, 13013-13018 (1986).
17. Tapley, P. M. & Murray, A. W. *Biochem. biophys. Res. Commun.* **118**, 835-841 (1984).
18. Witt, J. J. & Roskoski, R. Jr *Analyt. Biochem.* **66**, 253-258 (1975).
19. Lowry, O. H., Rosenbrough, N. J., Farr, A. L. & Randall, R. J. *J. biol. Chem.* **193**, 265-275 (1951).
20. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
21. Burnette, W. N. *Analyt. Biochem.* **112**, 195-203 (1981).
22. Wolf, M., Cuatrecasas, P. & Sahyoun, N. *J. biol. Chem.* **260**, 15718-15722 (1985).

Optical trapping and manipulation of single cells using infrared laser beams

A. Ashkin*, J. M. Dziedzic* & T. Yamane†

* AT&T Bell Laboratories, Holmdel, New Jersey 07733, USA

† AT&T Bell Laboratories, Murray Hill, New Jersey 07974, USA

Use of optical traps for the manipulation of biological particles was recently proposed, and initial observations of laser trapping of bacteria and viruses with visible argon-laser light were reported¹. We report here the use of infrared (IR) light to make much improved laser traps with significantly less optical damage to a variety of living cells. Using IR light we have observed the reproduction of *Escherichia coli* within optical traps at power levels sufficient to give manipulation at velocities up to $\sim 500 \mu\text{m s}^{-1}$. Reproduction of yeast cells by budding was also achieved in IR traps capable of manipulating individual cells and clumps of cells at velocities of $\sim 100 \mu\text{m s}^{-1}$. Damage-free trapping and manipulation of suspensions of red blood cells of humans and of organelles located within individual living cells of spirogyra was also achieved, largely as a result of the reduced absorption of haemoglobin and chlorophyll in the IR. Trapping of many types of small protozoa and manipulation of organelles within protozoa is also possible. The manipulative capabilities of optical techniques were exploited in experiments showing separation of individual bacteria from one sample and their introduction into another sample. Optical orientation of individual bacterial cells in space was also achieved using a pair of laser-beam traps. These new manipulative techniques using IR light are capable of producing large forces under damage-free conditions and improve the prospects for wider use of optical manipulation techniques in microbiology.

The trapping and manipulation of dielectric particles by laser light is based on forces due to radiation pressure^{2,3}. In the so-called single-beam gradient force traps used here^{1,4,5} the dominant force component is the gradient force which pushes dielectric particles into the high-intensity region of a highly focused light beam. For physical particles such traps have been shown to operate over a range of particle sizes from tens of microns down to hundreds of Å⁵ and also for individual atoms⁶.

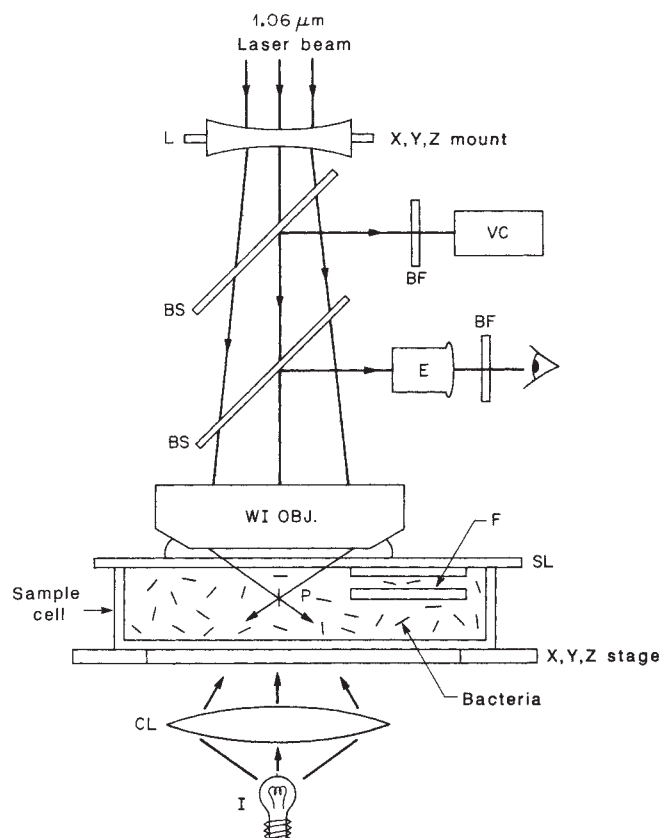


Fig. 1 Combined high-resolution optical microscope and 1.06 μm infra-red laser trap for observing, manipulating and separating bacteria and other organisms.

For operation with micron-sized biological particles the gradient trap was introduced into a high-resolution optical microscope where it can be used to manipulate particles within a sample while being viewed under high magnification¹.

In these experiments a 1.06 μm Nd:YAG (neodymium-doped yttrium aluminium garnet) laser trapping beam was coupled into a conventional microscope as shown in Fig. 1, with a beam splitter, BS, and focused with an adjustable external lens, L,

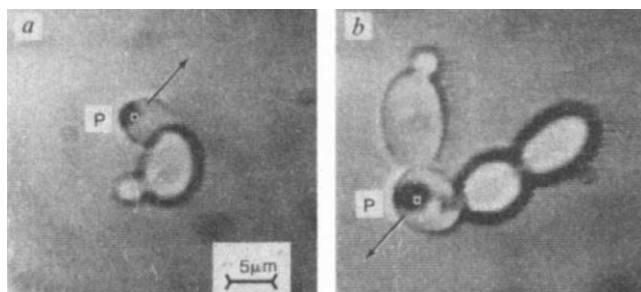


Fig. 2 The division of yeast cells in the IR trap. An original clump of two cells increases to four cells and then to six cells as shown in *a* and *b* after a total elapsed time of about 3 h.

into the viewing plane at point P. The high-numerical-aperture water-immersion objective (WI-OBJ) of the microscope (numerical aperture = 1.25) functions in the IR to form the single-beam gradient trap at P and in the visible with the high-numerical-aperture condenser lens, CL, and illuminator I to give high-resolution viewing. IR-blocking filters, BF placed in front of the eyepiece, E, and video camera, VC, serve to isolate the viewing optics from back reflections from the sample cell. The optical laser trap can be moved about within the $1\text{ cm} \times 3\text{ cm}$ $100\text{ }\mu\text{m}$ thick sample cell by moving either the external lens on its X, Y, Z mount or by moving the X, Y, Z microscope stage.

To observe the reproduction of *E. coli* within the laser trap, an individual bacterium was captured and lifted up from among a collection of bacteria that had settled at the bottom of the sample to a clear region, where it was monitored continuously. As individual rod-like bacteria line themselves up vertically along the trapping-beam axis, the bacteria were released momentarily every 10 min to determine better their size. In a run of almost 5 h about 2.5 life cycles were observed with all four of the resulting offspring remaining in the trap. The ability to reproduce within the laser trap clearly demonstrates that the bacterium was not damaged by the IR beam.

Separation of bacteria from one sample and their introduction into another was accomplished by first manipulating them into a liquid-filled hollow-glass fibre, F, with an inside diameter of

$14\text{ }\mu\text{m}$ and a length of 5 mm, which was attached to the lid, SL, of the sample cell. The lid and fibre were then removed, washed externally and blown dry, without disturbing the bacteria inside the small core of the liquid-filled fibre. The lid and fibre were later placed on top of another sample cell into which the bacteria could be transferred using the optical trap. Reproduction of *E. coli* while being held in the light trap inside the confines of the hollow fibre, was also observed.

In the above reproduction experiments no evidence of damage to *E. coli* was seen up to the maximum power available at $1.06\text{ }\mu\text{m}$ of 80 mW, as measured at the sample. Other more motile types of bacteria were also captured at 80 mW with no evidence of optical damage as indicated by the absence of changes in motility after many minutes of trapping. At this IR-power level it was possible to drag trapped bacteria at velocities up to $\sim 500\text{ }\mu\text{m s}^{-1}$ through the fluid. In previous experiments¹ with $0.51\text{ }\mu\text{m}$ visible laser light the threshold for damage to *E. coli* and other more motile bacteria occurred at considerably lower intensities and lower manipulation velocities. The manipulation velocity at the damage threshold is about an order of magnitude higher in the $1.06\text{ }\mu\text{m}$ IR traps than in the $0.51\text{ }\mu\text{m}$ visible traps. The trapping force corresponding to a $500\text{ }\mu\text{m s}^{-1}$ drag velocity is about 800 mg assuming a bacterium of $1\text{ }\mu\text{m}$ diameter and unit density. This is sufficient to capture most motile bacteria.

To demonstrate the ability to orient biological particles in space we introduced a second independently adjustable trapping beam into the sample cell. This is possible through another microscope objective at the bottom of the sample cell which functions as both an IR trapping lens and a visible-condenser lens or through the same trapping and viewing objective at the top using beam splitters.⁷ With two beams we were able to hold a rod-like bacterium, for example, at each of its ends and orient it at will.

Optical trapping of yeast cells (*Saccharomyces cerevisiae*), was achieved over a range of powers of about 5–80 mW. Individual yeast cells held in the laser trap were observed to bud into clumps of as many as eight connected cells; each achieving a size up to $\sim 5\text{--}10\text{ }\mu\text{m}$ in diameter over a time span of about 5 h. It was possible to manipulate individual cells at velocities up to $\sim 100\text{ }\mu\text{m s}^{-1}$ with a power of 80 mW. Manipulation at velocities of $\sim 30\text{ }\mu\text{m s}^{-1}$ was also possible holding any one of the cells of a single large clump. By comparison severe optical

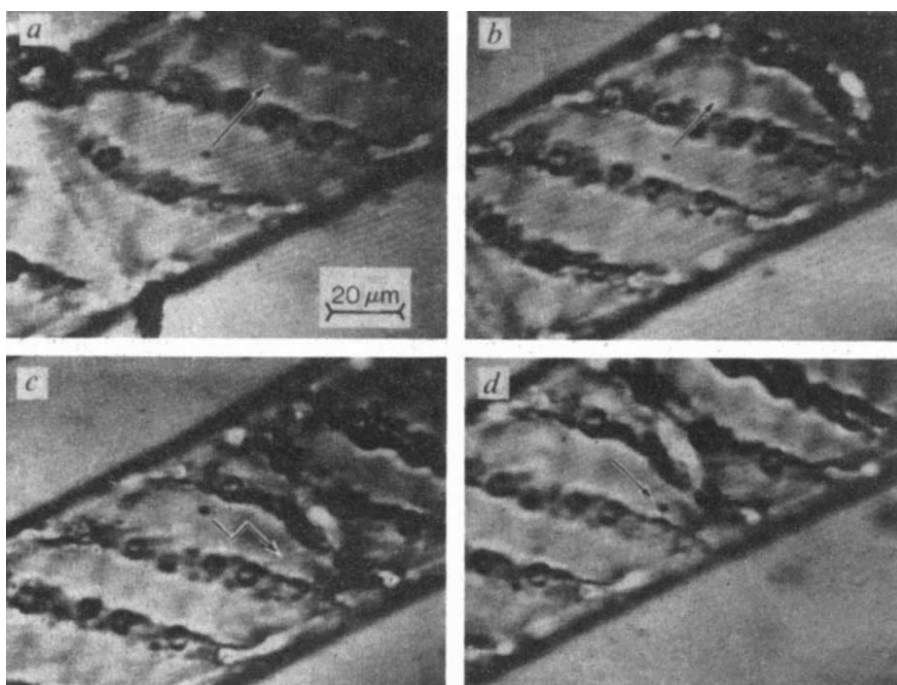


Fig. 3 *a–d*, The manipulation of a clump of particles within the cytoplasm of a cell of a spirogyra along the path of the arrows, by moving the microscope stage.

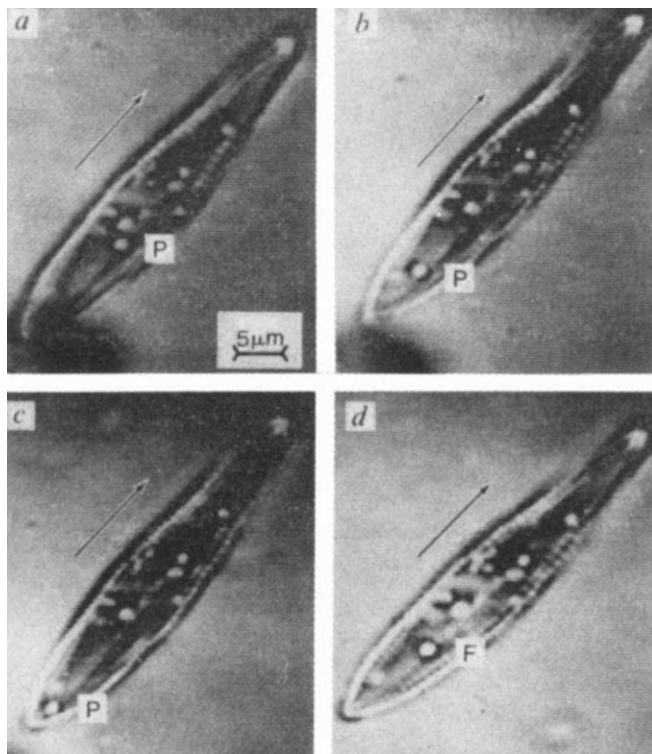


Fig. 4 *a-d*, The manipulation of an optically trapped organelle within the interior of a protozoan as it moves along freely on a glass surface in the direction of the arrow. In photographs *a-c*, an organelle, originally trapped at P (*a*), is being dragged by the advancing protozoan to the rear of the cell (*c*). In (*d*) the organelle pulls free of the trap and snaps back to its final position at F.

damage to yeast was seen after a few minutes using visible-light traps at $0.5145\ \mu\text{m}$ at lower force manipulation velocities and lower power levels ($\sim 20\ \text{mW}$). Figure 2 illustrates the reproduction of yeast cells in the IR trap. The first photograph (*a*), taken after about a half-hour in the trap, shows the increase of an original clump of two yeast cells to a clump of three with an incipient fourth bud appearing near the location of the trap at point P. The clump is being dragged along through the water of the cell in the direction of the arrow, by moving the microscope stage. This motion creates viscous drag that causes the clump, which at rest hangs vertically, to tip up into the viewing plane for better visibility. After $\sim 3\ \text{h}$ the clump has grown to six cells as shown in the second photograph (*b*).

Red blood cells (erythrocytes) of humans were trapped and manipulated with no apparent change in flexibility or appearance with powers of 4–40 mW in the IR. The extreme flexibility of these cells was evident by the ease with which their shape was distorted by optically generated forces. Trapping occurs most readily on the thick doughnut-shaped rim of the cell. Manipulation at a velocity of $\sim 100\ \mu\text{m s}^{-1}$ was observed at a power of 40 mW. At powers of 80 mW some loss of flexibility was seen after trapping for tens of minutes. With visible argon-laser light catastrophic damage occurred with only a few milliwatts of power and low force levels.

In experiments on individual living cells of *spirogyra*, we were able to scan a focused 80 mW beam over the entire cell with no sign of damage. When the chloroplasts or pyrenoid bodies were irradiated there were no changes in appearance or changes in the nearby cytoplasmic streaming. We were, however, able to trap individual particles or collections of particles of about a micron or less, which are characteristically seen streaming along

in well-defined channels in the cytoplasm near the cell walls. If manipulated into nearby interior regions of the cell vacuole, for example, the particles appear to drag part of the channels and cytoplasm with them and when released snap back as if on elastic bands. If some particles are freed totally from the channel, they can be manipulated about anywhere within the cell and used as probes for tracing out the internal structure of the cell. The cell appears to be sufficiently transparent and the trap strong enough to allow manipulation of particles under and around localized obstacles within the cell. Figure 3 shows a sequence of photographs illustrating manipulation, within the cytoplasmic layer of a cell of a *spirogyra*, of a small clump of captured steaming particles. The particles were moved along the path indicated by the arrows in Fig. 3*a-d*, taking them past the pyrenoid bodies and chloroplasts situated near the top cell wall of the *spirogyra*. Damage to *spirogyra* occurred within the chloroplasts using a visible-argon-laser light trap at powers of a few mW, probably due to chlorophyll absorption. The ability to manipulate organelles within the interior of a living cell without damaging the cell wall is probably unique to the optical-manipulation technique.

Experiments were also performed on the trapping of protozoa and of organelles within protozoa. Little evidence of damage due to the light was seen at powers up to $\sim 160\ \text{mW}$, as measured at the sample. Trapping was observed for free-swimming small motile protozoa and even varieties of surface-adhering protozoa that were occasionally pried free of a surface. In one instance an amoeba was successfully trapped free of any contacting surface and manoeuvred about as it continuously changed its shape. It was finally redeposited on a glass surface apparently undamaged. Figure 4 shows an example of manipulation of an organelle within the interior of a fairly large protozoan as the organism moves along freely on the bottom glass surface of the sample cell in the direction of the arrow. In Fig. 4*a*, the laser trap has just been manipulated to point P where it captures the prominent circular particle located at that position within the interior of the protozoan. As the protozoan advances, the trapped particle at P is pulled back toward the rear of the moving cell as shown in Fig. 4*b*. In Fig. 4*c* the restrained particle is seen colliding with the cell wall at the rear of the advancing protozoan. Subsequently, as seen in Fig. 4*d* the protozoan pulls the particle free of the trap and the released particle snaps back to a final position F, part of the way back to its original position in Fig. 4*a*. Observations such as these are clearly giving information on the viscosity and elastic properties of the cytoplasm in the region of the trapped organelle. It was also possible to manipulate particles within stationary protozoa by simply moving the trap.

Apart from possible reductions in absorption of the particles, IR traps have the further advantage over visible traps of a fourfold reduction in intensity due to the larger focal spot size, without a reduction in force, provided the local particle curvature does not vary significantly over the focal spot. Another factor in use of IR for trapping, is the temperature rise due to absorption of the water of the sample cell itself. At a power of 80 mW, assuming an absorption coefficient a of $\sim 0.1\ \text{cm}^{-1}$, the temperature rise is estimated to be several degrees Centigrade. At higher powers this effect will eventually cause damage to the trapped particle and be a limiting factor even if additional particle absorption does not occur.

Received 21 July; accepted 5 November 1987.

1. Ashkin, A. & Dziedzic, J. M. *Science* **235**, 1517–1520 (1987).
2. Ashkin, A. *Phys. Rev. Lett.* **24**, 156–159 (1970).
3. Ashkin, A. *Science* **210**, 1081–1088 (1980).
4. Ashkin, A. *Phys. Rev. Lett.* **40**, 729–732 (1978).
5. Ashkin, A., Dziedzic, J. M., Bjorkholm, J. E. & Chu, S. *Opt. Lett.* **11**, 288–290 (1986).
6. Chu, S., Bjorkholm, J. E., Ashkin, A. & Cable, A. E. *Phys. Rev. Lett.* **57**, 314–317 (1986).
7. Ashkin, A. & Dziedzic, J. M. *Appl. Opt.* **19**, 660–668 (1980).

Structure and mobility of electron gain and loss centres in proteins

George D. D. Jones, Jeremy S. Lea,
Martyn C. R. Symons* & Fatai A. Taiwo

Department of Chemistry, University of Leicester,
University Road, Leicester LE1 7RH, UK

Radiation damage to proteins is a topic of intense interest to those involved in radiation effects in biology, and also to those involved in radiotherapy. Although it has been widely studied, fundamental processes in protein damage are very hard to specify because of the complexity of the final damage products. But the non-invasive technique of electron-spin resonance is ideally suited to the task of detecting and identifying the primary and secondary products as these are expected to contain unpaired electrons (that is, free-radicals) and such species are uniquely detected by this sensitive form of spectroscopy. Our present study shows that a major radical species formed by electron loss in a range of proteins is the backbone amido radical, $-N^{\bullet}(CO)-$, characterized by hyperfine coupling to one ^{14}N nucleus. These centres are efficiently trapped in proteins at low temperatures. In contrast, the expected backbone electron-capture centres, $-NH(CO^{\bullet})-$, are not readily trapped and electron transfer occurs until the ejected electron is trapped by some electrophilic centre. Such electron mobility was in fact established in our previous work on oxyhaemoglobin ($FeO_2 \rightarrow FeO_2^{\bullet}$), superoxide dismutase ($Cu(II) \rightarrow Cu(I)$) haemocyanin ($Cu(II)O_2Cu(I) \rightarrow Cu(I)O_2Cu(II)$) and various proteins containing S—S bonds ($-S-S- \rightarrow -S-S^{\bullet}-$) (refs 1–4 respectively). This is strongly supported by our observation that such electrons are captured by DNA molecules, giving $T^{\bullet}$ centres, when nucleohistones are irradiated⁵, and that $Fe(CN)_6^{3-}$ ions readily scavenge such electrons from proteins which are devoid of highly electrophilic centres.

In previous studies it has been supposed that, on electron loss, either carbon-centred radicals



or specific side-group cations such as $Trp^{+\bullet}$ or $Tyr^{+\bullet}$ are formed^{6,7}. But in our view, the former must be secondary species and the latter will only dominate if the primary hole centres are mobile. Our expectation is that primary electron loss should give radical cations whose semioccupied molecular orbitals (SOMOs) are largely centred on nitrogen, as is the case for simple amides⁸. We have therefore studied the central 'free-spin' ($g=2$) regions of a range of irradiated proteins and certain polypeptides in frozen aqueous solutions (at 77 K). For systems containing electron-scavenging centres (metmyoglobin, methaemoglobin, oxymyoglobin, oxyhaemoglobin, $Cu(II)$ superoxide dismutase and various proteins containing S—S linkages) electron-spin resonance (ESR) signals which are assigned mainly to $>N$ -centred radicals (amide radicals) dominated. For other proteins (the five histones, bovine-serum albumin, collagen and others, together with certain polypeptides) when $Fe(CN)_6^{3-}$ ions were added as electron scavengers, similar ESR spectra were obtained at 77 K. For illustrative purposes, we describe our results for oxy- and methaemoglobin. Here the reactions $Fe(III) + e^- \rightarrow Fe(II)$ and $(FeO_2) + e^- \rightarrow (FeO_2)^{\bullet}$ are of high efficiency, as monitored by loss of the $Fe(III)$ signal at $g=6$, or gain of the signal for $(FeO_2)^{\bullet}$ units¹.

A typical spectrum is shown in Fig. 1. If, as we anticipate, the primary cations ($-CONH^+-CHR-$) readily deprotonate *via* their hydrogen bonds to give ($-CON-CHR-$) radicals, the following ESR features are expected by comparison with a wide range of nitrogen-centred π radicals. (1) Relatively weak $M(I)(^{14}N) = \pm 1$ 'parallel' features ($M(I)$ is the magnetic nuclear

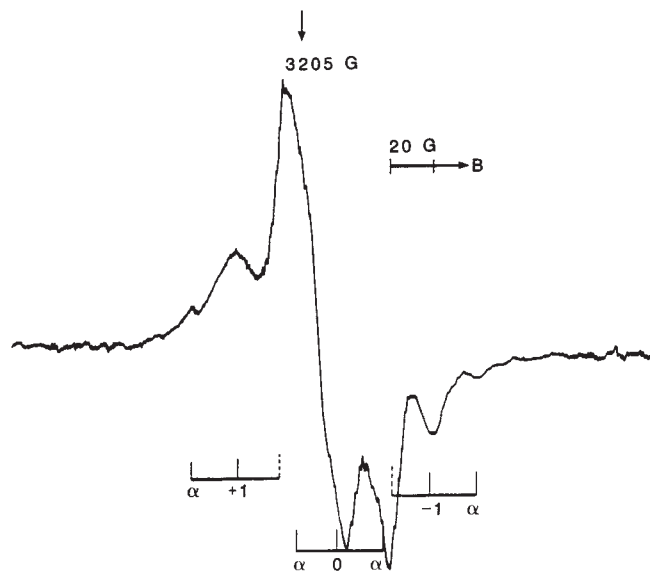


Fig. 1 First-derivative-ESR spectrum for methaemoglobin after exposure to ^{60}Co γ rays at 77 K and annealing to ~ 140 K to remove signals from $\cdot OH$ radicals in ice crystallites, showing features assigned to electron-loss centres. (Ejected electrons were captured at $Fe(III)$, as indicated by a marked decrease in the $Fe(III)$ feature at $g=6$.) The main species [$M(I)(^{14}N) = +1, 0, -1$] has $A_{||} = 43$ G, $A_{\perp} \approx 0$, and no resolved proton splitting: the other species (α) has similar ^{14}N parameters with $A(^1H) \approx 38$ G. (The $|+1, -\frac{1}{2}\rangle$ and $|+1, +\frac{1}{2}\rangle$ lines for α , indicated by dashed lines, are concealed beneath the main spectrum.)

spin quantum number) together with an intense nearly isotropic $M(I) = 0$ line. (2) Each of these should split into nearly isotropic doublets by coupling to the unique β protons of the $-CHR-$ unit attached to nitrogen. The difficulty with this prediction is that the magnitude of the coupling to β protons is a function of the angle θ defined in Fig. 1, and could take any value from ~ 50 G to zero, depending on this angle. For values of θ close to 90° we expect to detect unsplit features of significant intensity but unless there are a large number of units with nearly equal values for θ the remainder should be extensively broadened over the available range. Fortunately, as can be judged from Ramachandran plots for many proteins⁹, values of θ do tend to cluster and, in particular, for α -helix units, values close to 90° are expected.

Results for irradiated oxy- and methaemoglobin are in satisfactory agreement with these predictions (Fig. 1). For these proteins, electrons are efficiently captured by the iron units, so that the $g=2$ region of the spectra should be dominated by the electron-loss centres. The results show that the major species has a large ^{14}N -parallel splitting (~ 43 G) but a zero or small β -proton coupling (0 ± 5 G). This accords well with expectation. However, the value calculated for radicals in the α helix is 38 G. Outer features and a central ($M(I) = 0$) doublet with an absorbance of ($A(^1H)$) ~ 38 G can also be seen. As indicated, this splitting accommodates the outer parallel lines given that, as expected, $A_{||}(^{14}N)$ remains ~ 43 G. The central lines correspond to $\theta \approx 30^\circ$ which is close to the upper limit allowed by the Ramachandran plot. Definite turning points are expected for this upper limit.

It seems that these centres are trapped rapidly, before significant migration can occur. For example, when deoxyhaemoglobin was irradiated [with $Fe(CN)_6^{3-}$ to scavenge the electrons], there was no appearance of signals at $g=6$, characteristic of $Fe(III)$ haem centres. Thus none of the electron-loss centres were able to migrate to iron, whereas for the met $Fe(III)$ form, the majority of ejected electrons were able to reach the iron centre before reacting with amide centres. Similarly, whereas disulphide anions ($RSSR^{\bullet-}$) are major centres in proteins con-

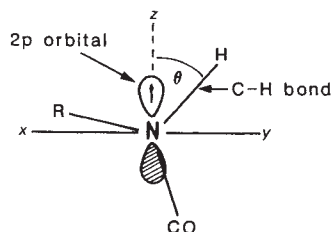


Fig. 2 View along the N—C bond with the amide unit in the $x-y$ plane, showing the angle θ between the $2p_z$ orbital on N and the C—H bond.

taining disulphide bridges, RSSR⁺ centres were not detected, and hence were not formed in significant yields.

Similar results were obtained for the other proteins. All these proteins contain α -helical regions and all gave clear parallel features with $A(^1\text{H}) \approx 38\text{G}$ for amide radicals formed in the α -helical regions. Other features corresponding to specific ^1H splittings were also found, which can be assigned to radicals in less highly organized regions of the protein. As yet, we have no clear information regarding the behaviour of electron-loss centres in β -pleat regions of proteins.

We conclude that for the proteins studied, electrons ejected by ionizing radiation have significant mobility and are efficiently

captured at electrophilic sites, but the hole-centres originating primarily from the amide backbone are rapidly trapped by loss of the N—H proton, and are well characterized by ESR spectroscopy. A range of such radicals are detected in any protein because of the variable β -proton coupling; those formed in α -helical regions being well defined with $A(^1\text{H}) \approx 38\text{G}$. It seems that electron-loss from the peptide backbone dominates. But additional unresolved central ESR features suggest the presence of delocalized organic radicals such as (Tyr)⁺ or (Trp)⁺ in some cases in relatively low yields. In most work, oxygen was removed. But the same radicals were formed initially in the presence of oxygen.

Received 12 August; accepted 4 November 1987.

1. Symons, M. C. R. & Petersen, R. L. *Proc. R. Soc. B* **201**, 285 (1978); *Biochim. biophys. Acta* **537**, 70 (1978).
2. Symons, M. C. R. & Stephenson, J. M. *J. chem. Soc., Faraday Trans. 1* **79**, 2983-2996 (1983).
3. Symons, M. C. R. & Petersen, R. L. *Biochim. biophys. Acta* **535**, 247-259 (1987).
4. Rao, D. N. R., Symons, M. C. R. & Stephenson, J. M. *J. chem. Soc., Perkin Trans. 2*, 727-733 (1983).
5. Cullis, P. M., Jones, G. D. D., Lea, J. S. & Symons, M. C. R. *J. chem. Soc., Chem. Commun.* (submitted).
6. Sevilla, M. D., D'Arcy, J. B. & Morehouse, K. M. *J. phys. Chem.* **83**, 2887-2895 (1979).
7. Liming, F. G. & Gordy, W. *Proc. natn. Acad. Sci. U.S.A.*, **60**, 795-806 (1968).
8. Rao, D. N. R. & Symons, M. C. R. *Chem. Phys. Lett.* **93**, 485-489 (1982); Eastland, G. W., Rao, D. N. R. & Symons, M. C. R. *J. chem. Soc., Faraday Trans. 1* **82**, 2833-2845 (1986).
9. Schulz, G. E. & Schirmer, R.-H. *Principles of Protein Structure* (Springer, Heidelberg, 1979).

Electron transfer from protein to DNA in irradiated chromatin

Paul M. Cullis*, George D. D. Jones,
Martyn C. R. Symons* & Jeremy S. Lea

Department of Chemistry, University of Leicester, University Road,
Leicester LE1 7RH, UK

Irradiation of dry or fully hydrated frozen DNA systems (conditions of direct damage) has been shown by electron-spin resonance spectroscopy to give rise to electron-gain centres localized on thymine (T⁻) and electron loss centres ('holes') localized on guanine (G⁺) with approximately equal yields¹⁻³. Our parallel studies on the development of both single- and double-strand breaks under comparable conditions provide good evidence that these radical centres are the precursors to such damage³⁻⁵, and we and others have argued that this may be of relevance to the damage pathways *in vivo*. We now report evidence that when DNA is complexed to proteins as it is in the nuclei of eukaryotes, electron transfer from the histone to DNA is facile, leading to a significant increase in the yield of electron-gain centres in DNA as judged from their electron-spin resonance spectra. In contrast 'holes' generated in the protein are trapped and do not lead to any detectable increase in the yields of G⁺.

There is indirect evidence that on irradiation of DNA:histone complexes, damage is in some way transferred from the protein to DNA^{6,7}. Furthermore, a study of strand breaks conducted on DNA:histone complexes suggests that the number of single-strand breaks is enhanced relative to uncomplexed DNA, although no detailed mechanism for this was proposed⁸.

In this study we have used purified chromatin and stripped cell nuclei and have compared the electron-spin resonance (ESR) signals obtained therefrom with those obtained from the histone proteins and the DNA separately. Identical conditions were used (γ -irradiation at 77 K plus annealing to 130 K to remove ESR signals from hydroxyl radicals ($\cdot\text{OH}$) trapped in ice crystals³) so that the results could be compared quantitatively. If there were no interaction, the ESR spectra from chromatin would be just the sum of those obtained from equal concentrations of the separate components. Unfortunately, most of the centres have ESR features that overlap in the $g=2$ 'free spin' region (that is G⁺, T⁻, Hist⁺ and Hist⁻), but the proton-

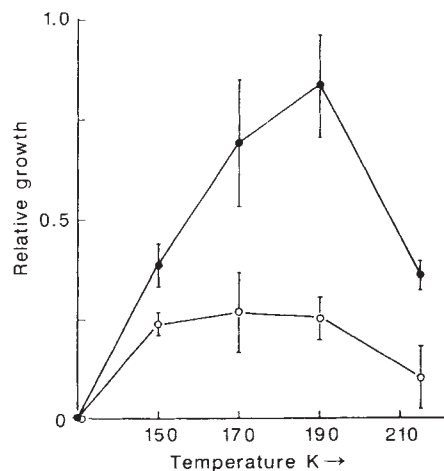


Fig. 1 Temperature-dependent growth of the 5-thymyl radical (TH·) ESR signal for chromatin (●) and DNA (○) (both at DNA concentrations of 15.4 mg ml⁻¹) after irradiation with ⁶⁰Co γ rays at 77 K estimated from the relative intensity of the seventh line of the TH· octet. The errors bars indicate a 99% confidence limit for duplicate samples.

ated form of T⁻, which builds up on annealing, with effectively complete conversion at $\sim 180\text{K}$ (Fig. 1) has a well-defined octet spectrum, the outer lines being completely separate from features for any of the other radicals. The results (Fig. 1) show an unambiguous gain in the concentration of the 5-thymyl radical, [TH·], and hence in [T⁻] when chromatin is used. This is true throughout the temperature range, although the maximum yield is achieved at a slightly higher temperature than for simple DNA systems.

This result accords with those for a range of proteins reported in the previous communication⁹. From this study, we expect electrons to be able to migrate to DNA, as observed, whereas electron-loss centres should remain trapped in the protein. It is much more difficult to assess changes in yields of Hist⁺ and G⁺ centres, but a series of computer subtractions involving spectra characteristic of each centre alone, support the concept that the yield of Hist⁺ and G⁺ remain unchanged, whereas that of Hist⁻ is greatly reduced and that of the T⁻ radical greatly enhanced. These spectral manipulations, shown in Fig. 2, are routine, and will be described in further detail elsewhere.

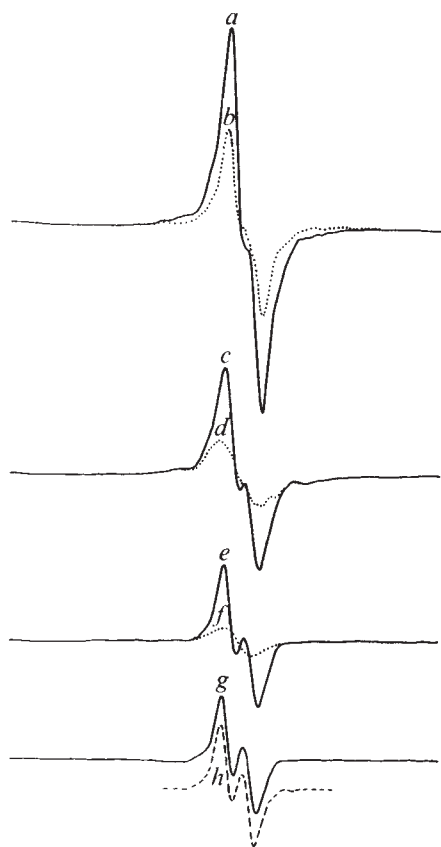


Fig. 2 The ESR spectra of: *a*, calf-thymus stripped cell nuclei after irradiation at 77 K and brief annealing to remove feature due to hydroxyl radicals; *b*, DNA at a concentration equivalent to the DNA concentration within the nuclei (19.4 mg ml^{-1}) treated under identical conditions; *c*, the result of subtraction of *b* from *a*; *d*, the result of subtraction of the histone radical-cation spectrum (*c*) from *c*; *e*, the result of subtraction of the histone radical-anion spectrum (*f*) from *e*; *h*, authentic $\text{T}^{\cdot-}$ for comparison.

From these results we conclude that, as with other protein systems³, the electrons are mobile and are able to migrate to sites of higher electron affinity, which in the case of the DNA:histone complex are the thymine bases. The electron-loss centres are not mobile, probably because these are centred on simple amide units and are trapped by loss of the N-H proton⁹. As we have already established that both $\text{T}^{\cdot-}$ and $\text{G}^{\cdot+}$ are precursors of strand breaks, as well as other forms of damage, the major increase in yield of $\text{T}^{\cdot-}$ is in our view highly significant. Firstly, it provides a clear explanation of the reported enhancement of single-strand breaks⁸ due to the presence of complexed histones and secondly, it suggests that nuclear DNA is more sensitive to ionizing radiation than isolated purified DNA.

We thank the Cancer Research Campaign for financial support and Dr D. Staynov for assistance in isolation of chromatin and nuclei.

Received 12 August; accepted 4 November 1987.

- Gregoli, S., Olast, M. & Bertinchamps, A. *J. Radiat. Res.* **89**, 238–249 (1982).
- Hüttermann, J. *et al. Discuss. Faraday Soc.* **67**, 135–145 (1984).
- Boon, P. J., Cullis, P. M., Symons, M. C. R. & Wren, B. W. *J. Chem. Soc. Perkin Trans. 2*, 1393–1398 (1984).
- Cullis, P. M. & Symons, M. C. R. *Radiat. Phys. Chem.* **27**, 93–101 (1986).
- Boon, P. J., Cullis, P. M., Symons, M. C. R. & Wren, B. W. *J. Chem. Soc. Perkin Trans. 2*, 1057–1062 (1985).
- Kuwabara, M. & Yoshii, G. *Biochim. biophys. Acta* **432**, 292–299 (1976).
- Stanger, H. K., Fielder, E. & Grillmaier, R. E. *Proc. Eighth Symp. Jülich, F.R.G.* 517–524; 671–683 (Commission of the European Communities, Radiation Protection, Microdosimetry, 1982).
- Lückle-Huhle, C., Braun, A. & Hagen, U. *Z. Naturf.* **25b**, 1264–1266 (1970).
- Jones, G. D. D., Lea, J. S., Symons, M. C. R. & Taiwo, F. A. *Nature* **330**, 772–773 (1987).

Copies of articles from this publication are now available from the UMI Article Clearinghouse.

For more information about the Clearinghouse, please fill out and mail back the coupon below.

UMI Article Clearinghouse

Yes! I would like to know more about UMI Article Clearinghouse. I am interested in electronic ordering through the following system(s):

- ☐ DIALOG/Dialorder ☐ ITT Dialcom
☐ OnTyme ☐ OCLC ILL Subsystem
☐ Other (please specify) _____
☐ I am interested in sending my order by mail.
☐ Please send me your current catalog and user instructions for the system(s) I checked above.

Name _____

Title _____

Institution/Company _____

Department _____

Address _____

City _____ State _____ Zip _____

Phone (____) _____

Mail to: University Microfilms International
300 North Zeeb Road, Box 91 Ann Arbor, MI 48106